Psychological Bulletin

EDITED BY

LYLE H. LANIER, VASSAR COLLEGE

CONTENTS

Psychological Bulletin

SPONTANEOUS ACTIVITY OF ANIMALS
A REVIEW OF THE LITERATURE SINCE 1929

J. DAVID REED

The Johns Hopkins University

Since the beginning of the century there has been a volume of work studying spontaneous activity of animals, especially the rat and monkey. The best review article is that of Shirley (97) in the *Psychological Bulletin* in 1929. Since that time the subject has not been reviewed comprehensively. Somewhat limited reviews appear in Munn (67), Morgan (66), and Gray (31). Other and even more circumscribed reviews include Richter (73, 74) reporting work done under his direction; Hoskins (40), describing some of the relationships between endocrines and activity; and Kreezer's (52) summary of methods for measuring activity in the rat. A review of diurnal rhythms by Welsh (119) in 1938 discusses much material not directly related to activity. Mettler (64) has reviewed and summarized studies on the effects of striatal injury in 1942.

This article does not attempt to cover work done prior to Shirley's review in THIS JOURNAL. Several of the most important references to work done prior to 1929 are included, but the emphasis has been almost entirely on later material.

THE CONCEPT AND MEASUREMENT OF SPONTANEOUS ACTIVITY

There are two points the reader should keep in mind as he proceeds through the paper. The first is a methodological issue. Much of the research to be reviewed depends on a general concept of spontaneous activity without regard to how the activity is measured. It will become evident in the course of the review that our concept of activity must be tied to the measure of it which we have used, for the results one gets with one measure of activity may be entirely reversed when a different measure is used. The second closely related point is a matter of terminology. Since the largest amount of work has used animals running inside a drum, it will simplify things if the term *activity*, without any qualification, always refers to running activity, not to other measures

of activity. Measures other than those in an activity-drum will always be clearly distinguished.

Method and Apparatus

Running Drums. Animals and human beings indulge in spontaneous activity. This observation has been quantified in many ways. The animal most frequently used in experimentation is the rat, whose activity is usually measured in what has been called an activity cage, but will henceforth be referred to as a drum or running drum. This device was first used by Stewart (110) and has been most adequately described by Slonaker (101, 102). It usually consists of two 10-13 inch circular boards mounted on a shaft and separated by a sheet of mesh wound around their periphery (86, 94). The rat runs inside the freely rotating drum, and a counter is attached to record the number of revolutions. Unfortunately, the usual system of measurement has shown only total activity, not activity as a function of time.

Recognizing this inadequacy, Skinner (100) used a Harvard work adder in conjunction with a kymograph to get a summative record whose slope is a constant measure of activity.

The drum has almost as many variations as there have been experimenters in activity. Stewart's 20-inch diameter drum and the 26-inch diameter drum used by Park and Woods (71) represent one extreme, while Shirley (94) used a 10-inch diameter. Results reported in terms of number of revolutions are obviously not comparable when the diameter of the drums is not the same. Furthermore, equating the running by expressing it in distance traversed is of questionable validity in view of Farris' statement that rats in larger wheels run farther than those in smaller wheels (24).

Depending upon the experiment, the rat may live entirely within the drum (94, 95, 101), have a separate living cage, or use supplemental diffuse activity cages (71). Since Richter (73) has shown that the number of revolutions of the drum is reduced when the rat has a choice of several things to do, the results of different experimenters may not be comparable.

The revolving drum has been the most extensively used laboratory instrument in investigating activity. Its physical variables have been discussed by Skinner (100) and Lacey (53). The reliability of the measures obtained is remarkable—Shirley (94) reports a rank-order correlations of .97 for five-day totals of activity, and a split-half r of .90. Beach's (4) figures are even higher (.98). Unfortunately a basic assumption in these results involves the equivalence of the measures. Lacey

(53) raises the very justifiable criticism that the measure may be showing only the consistency of the different drums. It is significant to note that in one case in which the animals were changed from one cage to another, the correlation reported was .80 (113). There are wide individual differences even between litter mates in normal rats with respect to running, some rats running 200 revolutions per day and others 20,000. The pattern of running is set up by the tenth day or not at all. After this time the individual differences are relatively constant.

The running drum has been used to indicate tension or motivation in the rat. Thus, Durrant (20) and Slonaker (106) have correlated running with sex drives. Geier and Tolman (28, 29) have used running behavior to indicate increase in tension in the rat.

Dorcus (16) devised a cage which moved slowly toward a goal object when the rat ran inside of it.

Tambour- or Spring-Mounted Cages. Another apparatus for measuring activity is that first used by Syzmanski (114, 115) which consisted of a spring-mounted cage attached to a lever recording system. The disadvantage of lack of damping has been somewhat overcome by tambour-mounted activity cages (73, 109). The three supporting tambours are joined to one tube and record every movement on a kymograph. Both these methods produce records according to time, but records which are difficult to treat quantitatively because no ready means of determining the total activity is available.

Hunt and Schlosberg (42, 43) counted the number of 5-minute active periods occurring in such a cage over varying intervals of time, and Irwin (44) recorded the number of active seconds per minute in newborn children. Wilbur (121) used a spring mounted cage connected to a Harvard work adder to obtain a summative record (of the activity of chicks) which is much easier to interpret.

Smith (109), measuring audiogenic and electrogenic convulsive activity, supported a cage from four pneumographs or by one large flexible hydron bellows. Oscillation could be reduced by means of a small vent.

Other animals have been used with appropriately modified cages to record activity. Monkeys have been fastened by a nine-inch chain to a 2.5-inch rod, so that movement caused the rod to advance a counter (84). A monkey-sized pneumatically-mounted activity cage has been used by Kennard, Massimy and Chevallier (51, 62).

Other Automatic Methods of Recording Activity. Another measure of activity has been suggested, incorporating a tilting box (7) in which the movement of the rat from one end of the box to the other advanced a counter. Claiming that the tilting motion would interfere with accurate

measurement of rat's activity, Siegel (98) utilized the animal's motion from one end of a 22×6 inch box to the other to break a photoelectric relay and thus advance a counter.

A horizontal turntable for exercising rats has also been used to record activity (21, 22). Since the distance the rat runs depends upon his proximity to the center of the turntable, it is probable that this device will not be popular in controlled experiments.

Curtis (15), working under Liddell, reports the use of a pedometer to record activity of the sheep and the pig. Head-shaking in chickens has also been reported by Levy (58).

Observational Method of Quantifying Activity. An observational means of recording activity has been used by Hall (32), Beach (3), and Fredericson (25). Hall recorded the distance traversed by rats in a round open field eight feet in diameter. Beach noted which of 36 squares a rat entered upon in the ten minutes it was free in an area three feet square. Fredericson observed six classes of behavior indulged in by rats in a field two feet square.

Spontaneous Activity as a Behavior Category

Now let us take a moment to see what the methods just reviewed have to do with the concept of activity. Most of the authors tend to lump all manifestations of activity together and to pin one label on all of them—*activity*. This failure to distinguish types of activity in terms of its measure leads to a false concept of activity, for what data we have point to more than one type, or at least more than one aspect of activity.

There are, for example, wide individual differences in the running activity of rats but not nearly such wide differences in restless cage activity. Tainter (116) found that caffeine, metrazol, and picrotoxin had no effect on running but did increase behavior measured in a diffuse activity cage. Hunt and Schlosberg (43) found only 9% decrease in diffuse activity with castration instead of the 98% found by Hoskins (40) for running activity.

In light of these considerable differences, it seems logical that different terms should be used to distinguish the two devices and the behaviors which they measure. Throughout this paper, the author has attempted to distinguish between running activity in the rotating drum, on the one hand, and diffuse activity in a cage or stabilimeter on the other. As long as the terminology for these two distinct situations is the same, the notion will tend to persist that they are strictly comparable measures, which they are not.

HEREDITY AND AGE

Genetic Basis. Rundquist (89) by selective breeding has been able to get active and inactive strains of rats. The active strain is less easy to purify than the inactive strain. Selection for running produced strains in which there were measurable: (1) increases in number of successful matings, (2) increases in sizes of litters, and (3) decreases in the gestation period. The active rats also had a higher basal metabolic rate than the inactive strain. The selective breeding of these strains has been carried through 29 generations, with no change beyond the 12th (9, 30). Brody has concluded that the two strains differ with respect to a single gene which acts as a dominant in males and a recessive in females. This gene must act as an inhibitor, since none of the matings within the inactive strain produces active offspring, but on the other hand, active-strain matings produce individuals which vary from extreme inactivity to extreme activity. The genetic factors are somewhat obscured by environmental influences.

Age. Running activity of rats increases with age until the animals are about 80 days old, then is relatively constant until about 120 days, after which it gradually falls off till death (72, 95, 101). Richter (73) determined the amount of running, diffuse activity, and nest building as a function of age. The more active rats tend to have shorter lives than the inactive rats.

THE INTERNAL ENVIRONMENT

Nutrition

The daily running activity of the rat increases just prior to the normal time of feeding, even though the rat has been fed in what would otherwise be an inactive period (72, 103). This fact is probably explicable on the basis of hunger or metabolic changes connected with the daily 24-hour hunger rhythm initiated by constant feeding at a specific hour (120). Studies made by Richter (73) on generalized activity, however, show that the rat very probably has a two-hour hunger rhythm if he is allowed to have food constantly available.

If a rat is deprived of food for a period of time, its activity will tend to increase for as much as 96 hours. If deprived of food and water, it increases for 72 hours before it drops off (73, 117), probably due to weakness. Brobeck (8) showed that if food intake and environmental temperature are held constant, there is a negative correlation between activity and weight gain. Smith and Conger (107) varied the diet of rats by keeping the caloric value constant but changing the proportion

of fat or protein. They found that up to 56% of the caloric value of food
may come from fat without reduction in spontaneous activity. Fifty
per cent of animal protein, however, induces marked reduction in
running. Following ingestion of protein there is a marked metabolic
rise which is not present following ingestion of fats. This result extends
Slonaker's (105) finding that a diet of 14 to 18% protein yields maximal
spontan___ running ___ ___ ___ ___ ___ ___, fe___ ___ eight grams
meat d___ ___ ___ ___ ___ ___ ___ ___ ___ ___

A p___ ___ ___ ___ ___ ___ ___ ___ ___ ___ ___ ___
activity ___ ___ ___ ___ ___ ___ ___ ___ ___ ___ ___
activity ___

Whe___ ___ ___ ___ ___ ___ ___ ___ ___ ___ ___
tained ___ ___ ___ ___ ___ ___ ___ ___ ___ ___ ___
drinkin___ ___ ___ ___ ___ ___ ___ ___ ___ ___ ___

Rats ___ ___ ___ ___ ___ ___ ___ ___ ___ ___ ___
flavin sh___ ___ ___ ___ ___ ___ ___ ___ ___ ___ ___
of the f___ ___ ___ ___ ___ ___ ___ ___ ___ ___ ___
degree t___ ___ ___ ___ ___ ___ ___ ___ ___ ___ ___
of Vitam___ ___ ___ ___ ___ ___ ___ ___ ___ ___ ___
rat; aft___ ___ ___ ___ ___ ___ ___ ___ ___ ___ ___
before a___ ___ ___ ___ ___ ___ ___ ___ ___ ___ ___
drop in ___ ___ ___ ___ ___ ___ ___ ___ ___ ___ ___
limited ___ ___ ___ ___ ___ ___ ___ ___ ___ ___ ___
low thro___ ___ ___ ___ ___ ___ ___ ___ ___ ___ ___
not com___ ___ ___ ___ ___ ___ ___ ___ ___ ___ ___
deprivat___ ___ ___ ___ ___ ___ ___ ___ ___ ___ ___
sympto___ ___ ___ ___ ___ ___ ___ ___ ___ ___ ___
results ___ ___ ___ ___ ___ ___ ___ ___ ___ ___

In a___ ___ ___ ___ ___ ___ ___ ___ ___ ___ ___
young ___ ___ ___ ___ ___ ___ ___ ___ ___ ___ ___
recover___ ___ ___ ___ ___ ___ ___ ___ ___ ___ ___
deprive___ ___ ___ ___ ___ ___ ___ ___ ___ ___ ___
diet to ___ ___ ___ ___ ___ ___ ___ ___ ___ ___ ___
active a___ ___ ___ ___ ___ ___ ___ ___ ___ ___

Tha___ ___ ___ ___ ___ ___ ___ ___ ___ ___ ___
is demo___ ___ ___ ___ ___ ___ ___ ___ ___ ___ ___
show a ___ ___ ___ ___ ___ ___ ___ ___ ___ ___ ___
compan___ ___ ___ ___ ___ ___ ___ ___ ___ ___ ___
rats fed ___ ___ ___ ___ ___ ___ ___ ___ ___ ___

untary running. The reduction in running occurred long before the
appearance of impaired running ability.

Drugs

There has been interest in the effects of drugs on behavior as showing
perhaps quantifiable action similar to that qualitatively produced in
human beings. In general, analeptics stimulate activity (6), but have a
_____ of depressing effects with prolonged use. Thus cocaine,
_____, _____, and _____ increased running for single
____, but _____ of _____ ensued when the specific drug was ad-
_____ for several days (116, 123). Caffeine, metrazol, and picro-
____ increased _____ activity, but decreased restless activity as
measured in diffuse activity cages (116). In a subsequent paper, Shulte
and Tainter (117) have shown the differential temporal effects of the
administration of caffeine, metrazol, and picrotoxin on diffuse activity.

Running activity is increased by benzedrine sulphate and by Kola,
but there is a decrement in activity less before doubling time 24 hours
___ _____.

Phenobarbital administration (tranquilizers?) (59) to white rats
increased the nighttime running level, but had no effect on the daytime
level which is normally much lower even than in the nighttime. These
animals [text] ate, drank less, and gained less weight. Even after
cessation of phenobarbital, the experimental rats were less active
than the controls (59).

Other drugs used have been thyroxin and thiouracil. These are
discussed below.

Endocrines

Adrenals. In a series of experiments dealing with endocrines up to
1930, Richter (90) reports that entire removal of the adrenals leads to
___ of voluntary movement. This operation is very difficult to perform
completely, due to diffuse ectopic [text] tissue. Reduction in activ-
ity [text] closely related to surviving adrenal particles. In
a _____ of 33 adrenalectomized male Richter (90) was able to restore
normal activity by transplant of adrenal glands to the ovaries. Subse-
quent [text] on circadian endocrine activity [text]. Histological ex-
amination [text] cortex through cases but no medulla. Increase of
salt in diet caused [text] a reliable but partial increase in activity.
Svenson [text] that with the absence strength of muscles
[text] adrenalectomy is
lowered. [text] bilateral adrenalectomy and sympathectomy and

adrenal inac...tion, for during in the exercise concerned without run-
ning activity....ceptitly led to growth of the body with no regulation even
after that ti....Late Long (40) found an appreciable decrease and there a
decrease in ....locomotor activity although....differences in resistance in
behavior in t....normally inactive never decreased markedly.

Thyroid. ...Conclusive increase and maintenance of running resulting
from thyroid...sage. The result was increase found by Hoskins, Lind-
say (34) who...found 90% reduction only by administration of and re-
ported by H.........Reduced TSG was gained per day. The amount ob-
tained depen......upon the presence of thyroid tissue remaining, the fraction
small piece is...enough to sustain activity. Decreased animals ay...may
be re-activat...by subsequent therapy. However, if too much...thyroid adm...istered to non-group the activity in most cases reduced.
That reducti...n in metabolic rate of rats is not responsible for decrease
in activity wa...shown by increasing the metabolism of the thyroidectom-
and normal ...ts in electroparied (44). Conclusion in this manner by
thiouracil ...sulted in normal growth, however so decreases, increased in
the length of...estimated period, increase of substances that affect the
disruption of...Circadian patterns in ACTIVE (65). Experiments result in
active rats ha...d been concluded that rats be healthy while (38)

Pituitary. ...A portion of the pituitary basis to regulate normal in
running activity. Removal of sustained of front anterior lobe of the
hypophysis re...ned the diminution in activity...of over. If intravenous one
of seven did ...her experiment the structures of rats be reactive to the
original level...While the enlarged hypomania to the anterior removal
of the eye had...only a disturbing, stress affect (61).

When the ...tion of run-running is normal to the immediate, the
normal oestra...persistence in various period given. The longest rats
showed oestra...period persistent by the differing intervals at about
9.6, 11.2, 13....14.7, 16.4. Absence number was at multiproduct a
whole multip...of the sensing day elimination intervals the same rate.
This explana...is unsubstantial by the fact that continuous expos-
nated all cycl...(41).

The genet...mum hypothesis explains the normal per-function to di-
urnal rhythm...(208).

Liver. ...Nervous function of the liver is important in moderating
activity. Liga...ion of the bile duct caused some accumulation in the
animals in w....the rate removed and active. Accumulation of the
effect was wi....Comparison of liver acting normal immediate of
the bile duct,...observed activity interacting without the liver. Hepa-
tectomy was...compared by Thorn over time (71) who noted that the

...producing time the same length of time to regenerate

...

...The decrease in activity following... ...tectomy is variable... ...This reduction is attributed to the inability to consolidate methods for...

...There are several successful early differences in daily running activity. The female guinea pig shows a level of activity of about one or two small peaks(?) ...the peak of which is reached at oestrus ...Clearly changes occur even there are changes in the distribution of oscillations as a function of the time in the oestrus cycle (103). ...can also be abolished by bilateral ovariectomy (__, 87), when the activity in the uterus is decreased but... ...but it is restored immediately afterwards (8?). The presence of the uterus is not essential to this activity cycle, as it is clear by the fact that hysterectomy has no permanent effect on running activity (18, 19). A corresponding time takes place during oestrus. No activity of maternal level pregnancy and lactation are seen, and the female is again in oestrus soon... ...ified epithelial cells in vaginal skin. Introduction of... ...during receptivity leads to running periodicity, with about 4 day... ...marked reduction in running, followed by gradually increasing activity until the previous level is reached in three or four (100?)

The cyclical pre-treatment can also be induced in females or decreased in maternal state, by several means, all of which involve concentration in hormonal chemistry. Progesterone given... ...women seven or nine times the usual amounts resulted in doubling activity was found to restore behaviour as well as any level (106). Administration of oestrogen in spayed females was equivalent again half a similar effect about... ...the earlier discrepancy, however, made its behaviour patterns, reported by Shirley, is readily explained by the findings of Young and Hill (1__?). They were able to restore activity to levels preceding before ovariectomy by expert management of routine, very close control over of the hormone thus introduced. A comparatively amount of oestrus is necessary for the consolidation of running activity, but beyond that quantity, further administrating has cancels the activity over-shoots periodicity. These are similar to Shirley's results on males (101), however... Castrated males show greater cyclic activity than normal males (now). Shirley... males increase activity and lose weight following injection of an appropriate fraction of hormone preparations (critical 101).

...The placing increased the activity gradually much less than that of the female, and little can be said about it here. Complete castration before sexual maturation hardly alters activity. Fractional

castration by G... of the ...
1, 1½, 1¾, 2, led... the following ...
his controls: 20% ...
for regeneration ...
birth leads to ...
statement has be... ...
in running almo... ...
Richter ascribes... this ...
controls. It is p... ...
berg (42, 43) wo... ...
tivity cages an... ...
minute active p... ...
activity of both ...
ular grafts, prov...

The relation ...
A male rat in a ...
activity correlat... ...
nearby drums, th... ...
of the simple oe... ...
relation between ...
Columbia obstr... ...
Rundquist (89) f... ...
those bred for i... ...
require clarificat...

General Features

The rat displ... more running and more diffuse activity in the dark
than in light. Th... showing of a 24 hour periodicity is a more highly
stable. The da... periodic animal ... ...
rhythm of the ar... ... glass chamber in safe proximity ...
external conditio... ... and temperature (10, 39, 41, 42, 86). ...
a rhythm involv... ... ... of light 3:00 to 6:00, and four hours
of light, Hemmi... ... was always were unable to abolish the ...
rhythm even in ... ... light under these conditions. Since both (20), by
using six hours of light and six hours alternating with six hours of total
darkness, Brown ... was able to introduce a 12 hour rhythm in about
30 out of 32 rats... Less striking was [10] of this process, as shown by
the fact that the longest period during which it persisted in any one ...
was 37 days, in spite of the continual 12-hour cycle of light and heat.

The female rat shows a four-to-five-day oestrous cycle of activity

... can be somewhat altered by lighting ... that constant light leads to ... hours of activity and lowered daily ... ... 1973. However [20] also found ... of constant cornifi-... ... and vaginal ... in the mouse. Subsequently he ... that the effect could disappear ... necessary for this ... ... sleep [15, 12]. Females ... showed the same ... ... in day as those reared in constant darkness. Varying schedules ... had little effect on the running rate. The ... intensity of the ... while it increased the running ... demonstrated by John-son [18] ... showed that the amount of ... of maximal activity ... it also related to the amount of cage ... ... These results ... combined with some of ... ... the mouse ... ... has more spontaneous activity ... than the rat. Wallace ... was less in Illinois rats on some. Hunt and Schlos-berg [14] attempted to determine the source of the 24-hour activity rhythm, after attempting to find out whether the rats ... they came to the conclusion that it was not found ... the house. However the normal day and light period caused the rats to shift their period of maximal activity to about a peak.

In constant light, the house rat will show peaks of activity which associated with a daily rest period [44]. This became true of blinded rats. The peak ... activity readings came to the end of the 12-hour rest period. It has previously been shown ... that ... rats will reduce their activity slightly ... [47]. Moore ... ... degrees Fahren-heit into animals in cyclones bins.

Temperature

The effect of temperature on the ... activity of ... two-day-old rats was described by Hill [11]. These young rats are still poikilo-thermic so the effect of external temperature had more effect on the body temperature. It was found that, following through the Arrhenius equation would fit the curves relating body ... to activity, quies-cence, and frequency of movements. Thus, he concluded that activity is a function of ... .

Activity Following Confinement

The effect that activity in restricted movement exercise theories has ... been little documented. ... [9] ... ... running ... after a ... period of confinement, ... but after ten days of enforced activity, running decreases. Siegel (s... found an increase in the number of interruptions of a light beam following close confinement.

It is important to note that in neither case is the amount of activity under conditions of no confinement significantly different from the increase reported after confinement.

Skinner (100) has stated that "if any extensive activity is prohibited during part of a day, the remaining part shows a greater 'density' of activity per unit of time." Since he measured activity for a five to six hour period beginning at 3:00 a.m., it is possible that his conclusion is an artifact based upon the normal peak of the rat's running activity which occurs at about 2:00 a.m. After this time there is a gradual decrease in the amount run, a rhythm which will persist in constant darkness. Without further work substantiating the generality of the principle of greater activity following inactivity, it would be parsimonious to ascribe Skinner's results to the specific hours at which he recorded activity.

Miscellaneous Studies

Other studies using the running drum which might be mentioned are: the correlation between activity and errors in learning a maze is low (89, 96); while that between activity and time to traverse mazes is higher (68); rats given difficult discriminations run less than rats that are not made abnormal by difficult problems (60); a series of electroconvulsive shocks greatly reduced voluntary activity in rats (112).

NEURAL CONTROL OF SPONTANEOUS ACTIVITY

The search for a neural center which controls activity has led to contradictory evidence. There is complete agreement that frontal lesions do increase activity and that bilateral lesions are more effective than uni-lateral lesions. The hyperactivity usually takes two to three weeks to emerge. The animals on which most of the work has been done are rats, cats, and monkeys.

The first observation of the effect of cortical lesions on activity has been attributed to several clinicians and experimenters. In 1920 Lashley (57) quantified the change in activity of the rat by using a running drum and noted that only frontoparietal lesions both increased the number of hours of running and decreased the time spent in resting. Jacobsen (45, 46) noted increased restlessness and general activity following frontal destruction in the monkey, and Langworthy and Kolb (56) described the behavior of cats with heightened restlessness. Since 1937, considerable evidence has been gathered concerning the effect on spontaneous activity produced by lesions in the brain.

Rats

In the rat, the unilateral removal of the frontal pole did not appear to augment running activity. The inactive rats became hyperactive while the active rats did not increase their running relatively so much. Bilateral ablation of the frontal poles was much more effective in increasing running than the unilateral (3, 83). Beach (4) measured running for 30 days before and 50 days after electrolytic destruction of varying amounts of the corpus striatum. Activity increased postoperatively in one animal, decreased in two, and was unchanged in two others. "No relationship between the magnitude of lesion and effects on activity could be determined. In the rat, the striatum evidently does not exert a controlling effect upon running activity as measured in this experiment." On the other hand, Richter and Hines (84) found that monkeys with unilateral striatal lesions immediately had greatly increased activity, and Mettler (64) reports hyperactivity in cats following striatal lesions.

According to Mettler (63, 64) when the striatum is injured, hyperkinesia is the rule. He asserts that the striatum is an inhibitory mechanism; " . . . stimulation of it produces inhibition and removal of it engenders evidence of motor release. It stands on the one hand between the cortex and the final common path as part of the route through which the cortex may exert an inhibitory effect and, on the other hand, it operates between the thalamus and lower motor mechanisms in the automatic inhibition incident to 'unconscious activity'." If the cerebral cortex is totally removed, the decorticated animal does not exhibit incessant activity but shows an inability to initiate or inhibit movement suddenly (65).

Cats

Cats with bilateral one-stage removal of the rostral portions of the cerebral hemispheres were noted by Magoun and Ranson (59) to be almost continually walking about. Langworthy and Richter (55) recorded increase in activity from 27 to 61 units from unilateral operation and to 399 units for the bilateral removal of motor cortex, premotor cortex, and possibly a small tip of the corpus striatum in cats.

Monkeys

In monkeys, Richter and Hines (84) found that bilateral removal of areas 8, 10, 11, and 12 had little effect on activity, while that of 9 did (62). On the other hand, Kennard and Ectors (50) reported increased activity following removal of area 8 alone. These results are not con-

tradictory in view of the method of measurement of activity. Richter and Hines (**84**) attached the monkey by a chain to a short steel rod projecting from an axle. Movement of the monkey caused the rod to advance a counter. The method of recording activity used by Kennard *et al.* was a pneumatically mounted diffuse activity cage.

Generally the activity builds up in the course of two to three weeks following an operation. Ruch and Shenkin (**88**), however, report that lesions in area 13 (of Walker) consistently produce hyperactivity within the second post-operative day. Richter and Hines (**84**) also report such immediate hyperkinesia when the monkey has striatal lesions.

Kennard *et al.* (**51**) have stressed the visual role in hyperactivity in monkeys. "Hyperactivity is markedly affected by visual stimuli. It disappears in the dark or when the animals have been deprived of vision either by enucleation of the eyes of by bilateral lobectomy. Absence of auditory stimuli has not the same effect."

A decrease in activity was noted by Barris (**2**) following "bilateral one-stage removal of the rostral portions of the neo-cortex of cats." Kennard *et al.* reported that "hypermotility in monkeys and chimpanzees is related to lesions of the rostral portions of area 6 and to area 8" (**51**).

SUMMARY

The literature of the last twenty years concerning activity in animals has been reviewed. The methods, results and concepts of activity have been summarized and appraised.

Several methods have been in use:

1. The running drum: this apparatus yields high reliability for measures taken on a particular drum, but there are also inconsistencies from one drum to another and from one experimenter's design of drum to another.

2. The diffuse activity cage: a cage mounted on tambours or springs. The record which it gives varies widely from one cage to another.

3. Several miscellaneous mechanical and observational methods, which have not been extensively used.

There are considerable individual differences in running activity as well as variability of one animal's activity from time to time. The individual differences are due in part to heredity, but are somewhat complicated by environmental influences. Intra-animal variability can be ascribed to several factors: running increases during hunger and most deprivations, during darkness and cool periods, and during oestrus. In various kinds of endocrine imbalance or deficiency, there is usually a decrease in activity—a marked decrease in the running drum but only a small decrement in the diffuse activity cage.

Running activity and diffuse activity are sometimes affected in the same way, sometimes differentially. Both reach a maximum during the cool or dark part of a 24-hour cycle. Some of the analeptic drugs stimulate both kinds of activity, but other drugs may increase diffuse activity while decreasing running activity.

Injury to the brain affects activity. In particular, lesions of the frontal cortex heighten activity, and bilateral lesions cause a greater increase than unilateral injury. Still, it is not yet clear whether there is a specific activity center in the brain and, if so, where it is.

There is now a very large body of data concerning animal activity, but it needs further definition and interpretation. Particularly needed is a clarification of the concept of activity in relation to the method of measuring it. Most treatments of the subject tend to regard activity as a single entity. Yet, in some cases, where comparable measures of activity are available from different devices, running drum and diffuse activity cage, the results are not the same. Activity, it would then appear, does not constitute a single behavior category which can be measured with any instrument but must be considered, for the present at least, in terms of its method of measurement.

BIBLIOGRAPHY

1. BACQ, Z. M. The effects of abdominal sympathectomy, adrenal inactivation and removal of the stellate ganglia on the spontaneous activity of the albino rat. *Endocrinology*, 1931, **15**, 34–40.

2. BARRIS, R. W. Cataleptic symptoms following bilateral cortical lesions in cats. *Amer. J. Physiol.*, 1937, **119**, 213–220.

3. BEACH, F. A. Effects of brain lesions upon running activity in the male rat. *J. comp. Psychol.*, 1941, **31**, 145–179.

4. BEACH, F. A. Effects of lesions to corpus striatum upon spontaneous activity in the male rat. *J. Neurophysiol.*, 1941, **4**, 191–195.

5. BLOOMFIELD, A., & TAINTER, M. L. The effect of vitamin B deprivation on spontaneous activity of the rat. *J. Lab. clin. Med.*, 1943, **28**, 1680–1690.

6. BOUGHTON, L. L. The effect of life cycle therapeutic dosage administration of drugs to albino rats. II.

On activity, maze learning and relearning. *J. Amer. pharm. Ass.*, 1942, **31**, 240–244.

7. BOUSFIELD, W. A., & MOTE, F. A. The construction of a tilting activity cage. *J. exp. Psychol.*, 1943, **32**, 450–451.

8. BROBECK, J. R. Effects of variations in activity, food intake and environmental temperature on weight gain in the albino rat. *Amer. J. Physiol.*, 1945, **143**, 1–5.

9. BRODY, E. G. The genetic basis of spontaneous activity in the albino rat. *Comp. Psychol. Monogr.*, 1942, **17**, No. 5. Pp. 24.

10. BROWMAN, L. G. Light in its relation to activity and estrous rhythms in the albino rat. *J. exp. Zool.*, 1937, **75**, 375–388.

11. BROWMAN, L. G. The effect of bilateral optic enucleation on the voluntary muscular activity of the albino rat. *J. exp. Zool.*, 1942, **91**, 331–344.

12. BROWMAN, L. G. The effect of bilat-

eral optic enucleation upon the activity rhythms of the albino rat. *J. comp. Psychol.*, 1943, **36**, 33–46.

13. Browman, L. G. The effect of controlled temperatures upon the spontaneous activity rhythms of the albino rat. *J. exp. Zool.*, 1943, **94**, 477–489.

14. Browman, L. G. Modified spontaneous activity rhythms in rats. *Amer. J. Physiol.*, 1944, **142**, 633–637.

15. Curtis, Q. F. Diurnal variation in the free activity of sheep and pig. *Proc. Soc. exp. Biol. & Med.*, 1937, **35**, 566–567.

16. Dorcus, R. M. A new device for studying motivation in rats. *J. comp. Psychol.*, 1934, **18**, 149–151.

17. Dugal, L. P., & Ross, S. Effet de l'ablation partielle du foie sur l'activité spontanée du rat blanc. *Rev. Canad. Biol.*, 1943, **2**, 435–441.

18. Durrant, E. P. Studies on vigor. XI. Relation of hysterectomy to voluntary activity in the white rat. *Amer. J. Physiol.*, 1927, **82**, 14–18.

19. Durrant, E. P. Relation of hysterectomy of long standing to voluntary activity in the white rat. *Amer. J. Physiol.*, 1931, **97**, 519. (Abstr.)

20. Durrant, E. P. Influence of the female white rat in bodily activity of the male. *Amer. J. Physiol.*, 1935, **113**, 37. (Abstr.)

21. Farris, E. J., & Enguall, G. Turntable for exercising rats. *Science*, 1939, **90**, 144.

22. Farris, E. J. Apparatus for recording cyclical activity in the rat. *Anat. Rec.*, 1941, **81**, 357–361.

23. Farris, E. J. Leucopenia associated with normal estrous in the albino rat. *Anat. Rec.*, 1942, **82**, 147–151.

24. Farris, E. J. Breeding of the rat. Ch. 1 in J. Q. Griffith, Jr. and E. J. Farris (Eds.), *The rat in laboratory investigation.* Philadelphia: Lippincott, 1942. Pp. 1–17.

25. Fredericson, E. The theory of psychomotion as applied to a study of temperament. *J. comp. Psychol.*, 1946, **39**, 77–89.

26. Gans, H. M. Studies in vigor. XIII. The effect of early castration on the voluntary activity of male albino rats. *Endocrinology*, 1927, **11**, 141–144.

27. Gans, H. M. Studies on vigor. Effect of fractional castration on the voluntary activity of male albino rats. *Endocrinology*, 1927, **11**, 145–148.

28. Geier, F. M. The measurement of tension in the rat. A contribution to method. *J. comp. Psychol.*, 1942, **34**, 43–49.

29. Geier, F. M., & Tolman, E. C. Goal distance and restless activity. I. The goal gradient of restless activity. *J. comp. Psychol.*, 1943, **35**, 197–204.

30. Graves, E. A. The genetic basis of activity in the albino rat. *Psychol. Bull.*, 1937, **34**, 757–758. (Abstr.)

31. Gray, W. L. *The effects of forced activity on maze learning and the selection and consumption of food by rats.* Unpublished Ph.D. Thesis, Johns Hopkins Univ., 1933.

32. Hall, C. S. Emotional behavior in the rat. III. The relationship between emotionality and ambulatory activity. *J. comp. Psychol.*, 1936, **22**, 345–352.

33. Hall, V. E., & Lindsay, M. The effect of dinitrophenol on the spontaneous activity of the rat. *J. Pharm. exp. Therap.*, 1934, **51**, 430–434.

34. Hall, V. E., & Lindsay, M. The relation of the thyroid gland to the spontaneous activity of the rat. *Endocrinology*, 1938, **22**, 66–72.

35. Hausmann, M. F. The behavior of albino rats in choosing food and stimulants. *J. comp. Psychol.*, 1932, **13**, 279–309.

36. Heller, R. E. Spontaneous activity in male rats in relation to testis

hormone. *Endocrinology*, 1932, **16**, 626–632.

37. HEMMINGSEN, A. M., & KRARUP, N. B. Rhythmic diurnal variations in the oestrus phenomena of the rat and their susceptibility to light and dark. Det. Kgl. Danske Videnskabernes Selskab., *Biologiske Meddelelser*, 1937, **13**, 1–61.

38. HERRING, V. V., & BRODY, S. Growth and development. XLIII. Diurnal metabolic and activity rhythms. *Univ. Missouri Agri. Exp. Station Res. Bull.* 274, 1938 (not seen).

39. HITCHCOCK, F. A. The effect of low protein and protein-free diets and starvation on the voluntary activity of the albino rat. *Amer. J. Physiol.*, 1928, **84**, 410–416.

40. HOSKINS, R. G. Studies on vigor. XVI. Endocrine factors in vigor. *Endocrinology*, 1927, **11**, 97–105.

41. HOSKINS, R. G., & BEVIN, S. The effect of fractionated chorionic gonadotropic extract on spontaneous activity and weight of elderly male rats. *Endocrinology*, 1941, **27**, 927–931.

42. HUNT, J. McV., & SCHLOSBERG, H. General activity in the male white rat. *J. comp. Psychol.*, 1939, **28**, 23–38.

43. HUNT, J. McV., & SCHLOSBERG, H. The influence of illumination upon general activity in normal, blinded and castrated male white rats. *J. comp. Psychol.*, 1939, **28**, 285–298.

44. IRWIN, O. C. Effect of strong light on the body activity of newborns. *J. comp. Psychol.*, 1941, **32**, 233–236.

45. JACOBSEN, C. F. A study of cerebral function in learning. The frontal lobes. *J. comp. Neurol.*, 1931, **52**, 271–340.

46. JACOBSEN, C. F. Studies of cerebral function in primates. *Comp. Psychol. Monogr.*, 1936, **13**, No. 3. Pp. 68.

47. JACKWAY, I. Voluntary activity in the rat as related to the intake of whole yeast. *J. comp. Psychol.*, 1938, **26**, 157–162.

48. JOHNSON, M. S. Effect of continuous light on periodic spontaneous activity of white-footed mice. *J. exp. Zool.*, 1939, **82**, 315–328.

49. JONES, M. R. The effect of phenobarbital on food and water intake, activity level, and weight gain in the white rat. *J. comp. Psychol.*, 1943, **35**, 1–10.

50. KENNARD, MARGARET A., & ECTORS, L. Forced circling in monkey following lesions of the frontal lobes. *J. Neurophysiol.*, 1938, **1**, 45–54.

51. KENNARD, MARGARET A., SPENSER, SUSAN, & FOUNTAIN, G., JR. Hyperactivity in monkeys following lesions of the frontal lobes. *J. Neurophysiol.*, 1941, **4**, 512–524.

52. KREEZER, G. L. Technics for the investigation of psychological phenomena in the rat. Ch. 10 in J. Q. Griffith, Jr. and E. J. Farris (Eds.), *The rat in laboratory investigation.* Philadelphia: Lippincott, 1942. Pp. 199–273.

53. LACEY, O. L. A revised procedure for the calibration of the activity wheel. *Amer. J. Psychol.*, 1944, **57**, 412–420.

54. LACEY, O. L. The dependence of behavior disorders in the rat upon blood composition. *J. comp. Psychol.*, 1945, **38**, 277–284.

55. LANGWORTHY, O. R., & RICHTER, C. P. Increases in spontaneous activity aroused by frontal lobe lesions in cats. *Amer. J. Physiol.*, 1939, **126**, 158–161.

56. LANGWORTHY, O. R., & KOLB, L. C. The experimental production in the cat of a condition simulating pseudo-bulbar palsy. *Amer. J. Physiol.*, 1935, **111**, 571–577.

57. LASHLEY, K. S. Studies of cerebral function in learning. II. *Psychobiology*, 1920, **2**, 55–136.

58. LEVY, D. M. On the problems of

movement restraint, tics, stereotyped movements, hyperactivity. *Amer. J. Orthopsychiat.*, 1944, **14**, 644–671.

59. MAGOUN, H. W., & RANSON, S. W. The behavior of cats following bilateral removal of the rostral portion of the cerebral hemispheres. *J. Neurophysiol.*, 1938, **1**, 39–44.

60. MAIER, N. R. F., & WAPNER, S. Studies of abnormal behavior in the rat. *J. comp. Psychol.*, 1944, **37**, 151–158.

61. MANN, C. W. The effect of thiouracil upon the heart rate, estrous cycle and spontaneous activity of the white rat. *J. Psychol.*, 1945, **20**, 91–99.

62. MASSIMY, R., & CHEVALLIER, R. J. Les effets, chez le singe, de l'ablation préfrontal unilatérale; modifications de l'activité, du mode réactionnel et des réflexes. *C. R. Soc. Biol. Paris*, 1942, **136**, 103–106.

63. METTLER, F. A. Relation between pyramidal and extra-pyramidal function. *Res. Publ. Ass. nerv. ment. Dis.*, 1942, **21**, 150–227.

64. METTLER, F. A., & METTLER, C. C. The effects of striatal injury. *Brain*, 1942, **65**, 242–255.

65. METTLER, F. A., METTLER, C. C., & CULLER, E. A. Effects of total removal of cerebral cortex. *Arch. Neurol. Psychiat., Chicago*, 1935, **34**, 1238–1249.

66. MORGAN, C. T. *Physiological psychology.* New York: McGraw-Hill, 1943.

67. MUNN, N. L. *An introduction to animal psychology: the behavior of the rat.* New York: Houghton Mifflin, 1933. Pp. 50–78.

68. OMWAKE, L. The activity and learning of white rats. *J. comp. Psychol.*, 1933, **16**, 275–285.

69. OMWAKE, L. The influence of barbital on the activity and learning of white rats. *J. comp. Psychol.*, 1933, **16**, 317–325.

70. OSBORN, C. M. Spontaneous diurnal activity in a genetically hypopituitary animal, the dwarf rat. *Anat. Rec.*, 1940, **78**, Suppl. p. 137.

71. PARK, O., & WOODS, L. P. A modified Hemmingsen-Krarup mammalian activity recorder. *Proc. Soc. exp. Biol. Med.*, 1940, **43**, 366–370.

72. RICHTER, C. P. A behavioristic study of the activity of the rat. *Comp. Psychol. Monogr.*, 1922, **1**, 1–55.

73. RICHTER, C. P. Animal behavior and internal drives. *Quart. Rev. Biol.*, 1927, **2**, 307–343.

74. RICHTER, C. P. Symposium: Contributions of psychology to the understanding of problems of personality and behavior. IV. Biological foundations of personality differences. *Amer. J. Orthopsychiat.*, 1932, **2**, 345–354.

75. RICHTER, C. P. The role played by the thyroid gland in the production of gross body activity. *Endocrinology*, 1933, **17**, 73–87.

76. RICHTER, C. P. The effect of early gonadectomy on the gross body activity of rats. *Endocrinology*, 1933, **17**, 445–450.

77. RICHTER, C. P. Cyclical phenomena produced in rats by section of the pituitary stalk and their possible relation to pseudo-pregnancy. *Amer. J. Physiol.*, 1933, **106**, 80–89.

78. RICHTER, C. P. Pregnancy urine given by mouth to gonadectomized rats: its effect on spontaneous activity and on the reproductive tract. *Amer. J. Physiol.*, 1934, **110**, 499–512.

79. RICHTER, C. P. The spontaneous activity of adrenalectomized rats treated with replacement and other therapy. *Endocrinology*, 1936, **20**, 657–666.

80. RICHTER, C. P., & BENJAMIN, J. A., JR. Ligation of the common bile duct in the rat. *Arch. Path.*, 1934, **18**, 817–826.

81. RICHTER, C. P., & ECKERT, J. F. The effect of hypophyseal injection and implants on the activity of hypophysectomized rats. *Endocrinology*, 1937, 21, 481–488.

82. RICHTER, C. P., & HARTMAN, C. G. The effect of injection of amiotin on the spontaneous activity of gonadectomized rats. *Amer. J. Physiol.*, 1934, 108, 136–143.

83. RICHTER, C. P., & HAWKES, C. D. Increased spontaneous activity and food intake produced in rats by removal of frontal poles of the brain. *J. Neurol. Psychiat.*, *Chicago*, 1939, 2, 231–242.

84. RICHTER, C. P., & HINES, M. Increased spontaneous activity produced in monkeys by brain lesions. *Brain*, 1938, 61, 1–16.

85. RICHTER, C. P., & SCHMIDT, E. C. H., JR. Behavior and anatomical changes produced in rats by pancreatectomy. *Endocrinology*, 1939, 25, 698–706.

86. RICHTER, C. P., & WANG, G. H. New apparatus for measuring the spontaneous motility of animals. *J. Lab. Clin. Med.*, 1926, 12, 289–292.

87. RICHTER, C. P., & WISLOCKI, G. B. Activity studies on castrated male and female rats with testicular grafts, in correlation with histological studies of the grafts. *Amer. J. Physiol.*, 1928, 76, 651–660.

88. RUCH, T. C., & SHENKIN, H. A. The relation of area 13 on the orbital surface of the frontal lobe to hyperactivity and hyperphagia in monkeys. *J. Neurophysiol.*, 1943, 6, 349–360.

89. RUNDQUIST, E. A. Inheritance of spontaneous activity in rats. *J. comp. Psychol.*, 1933, 16, 415–438.

90. RUNDQUIST, E. A., & BELLIS, C. J. Respiratory metabolism of active and inactive rats. *Amer. J. Physiol.*, 1933, 106, 670–675.

91. SCHULTE, J. W., TAINTER, M. L., & DILLE, J. M. Comparison of different types of central stimulation from analeptics. *Proc. Soc. exper. Biol. Med.*, 1939, 42, 242–248.

92. SEARLE, L. V., & BROWN, C. W. Effect of subcutaneous injections of benzedrine sulphate on the activity of white rats. *J. exper. Psychol.*, 1938, 22, 480–490.

93. SEARLE, L. V., & BROWN, C. W. Effect of variation in the dose of benzedrine sulphate on the activity of white rats. *J. exp. Psychol.*, 1938, 22, 555–563.

94. SHIRLEY, MARY. Studies in activity. I. The consistency of the revolving drum method of measuring the activity of the rat. *J. comp. Psychol.*, 1928, 8, 23–38.

95. SHIRLEY, MARY. Studies in activity. II. Activity rhythms, age and activity, activity after rest. *J. comp. Psychol.*, 1928, 8, 159–186.

96. SHIRLEY, MARY. Studies in activity. IV. The relation of activity to maze learning and brain weight. *J. comp. Psychol.*, 1928, 8, 187–195.

97. SHIRLEY, MARY. Spontaneous activity. *Psychol. Bull.*, 1929, 26, 341–365.

98. SIEGEL, P. S. A simple electronic device for the measurement of the gross bodily activity of small animals. *J. Psychol.*, 1946, 21, 227–236.

99. SIEGEL, P. S. Activity level as a function of physically enforced inaction. *J. Psychol.*, 1946, 21, 285–291.

100. SKINNER, B. F. The measurement of "spontaneous activity." *J. gen. Psychol.*, 1933, 9, 3–24.

101. SLONAKER, J. R. The normal activity of the white rat of different ages. *J. comp. Neurol. Psychol.*, 1907, 17, 342–359.

102. SLONAKER, J. R. Description of apparatus for recording the activities of small mammals. *Anat. Rec.*, 1908, 2, 116–122.

103. SLONAKER, J. R. Analysis of daily

activity of the albino rat. *Amer. J. Physiol.*, 1925, **73**, 485–503.

104. SLONAKER, J. R. Pseudopregnancy in the albino rat. *Amer. J. Physiol.*, 1929, **89**, 406–416.

105. SLONAKER, J. R. Effect of different per cents of protein in the diet. II. Spontaneous activity. *Amer. J. Physiol.*, 1931, **96**, 557–561.

106. SLONAKER, J. R. Sex drive in rats. *Amer. J. Physiol.*, 1935, **112**, 176–181.

107. SMITH, E. A., & CONGER, R. M. Spontaneous activity in relation to diet in the albino rat. *Amer. J. Physiol.*, 1944, **142**, 663–665.

108. SMITH, P. K., & SMITH, A. H. The effect of a diet low in inorganic constituents on the voluntary activity of the albino rat. *Abstr. Proc. Amer. physiol. Soc.*, 1934.

109. SMITH, K. U. An accurate method of recording activity in animals. *J. gen. Psychol.*, 1942, **27**, 355–358.

110. STEWART, C. C. Variations in daily activity produced by alcohol and by changes in barometric pressure and diet, with a description of recording methods. *Amer. J. Physiol.*, 1898, **1**, 40–56.

111. STIER, T. J. B. "Spontaneous activity" of mice. *J. gen. Psychol.*, 1930, **4**, 67–101.

112. STONE, C. P. Effects of electro-convulsive shocks on daily activity of albino rats in revolving drums. *Proc. Soc. exp. Biol., N. Y.*, 1946, **61**, 150–151.

113. STONE, C. P., & BARKER, R. B. Spontaneous activity, direct and indirect measures of sexual drives in adult male rats. *Proc. Soc. exp. Biol., N. Y.*, 1934, **32**, 195–199.

114. SZYMANSKI, J. S. Eine methode zur Einversuchung der Ruhe und Aktivitätsperioden bei Tieren. *Arch. f. d. ges. Physiol.*, 1914, **158**, 343–385.

115. SYZMANSKI, J. S. Activität und Ruhe bei Tieren und Menschen. *Z. allg. Physiol.*, 1920, **18**, 105–162.

116. TAINTER, M. L. The effects of certain analeptic drugs on spontaneous running activity of the white rat. *J. comp. Psychol.*, 194*, **36**, 143–155.

117. WALD, G., & JACKSON, B. Activity and nutritional deprivation. *Proc. nat. Acad. Sci., Wash.*, 1944, **30**, 255–263.

118. WANG, G. H. The relation between spontaneous activity and estrous cycle in the white rat. *Comp. Psychol. Monogr.*, 1923, **2**, 1–27.

119. WELSH, J. H. Diurnal rhythms. *Quart. Rev. Biol.*, 1938, **13**, 123–139.

120. WERTHESSEN, N. T. The significance of sub-normal respiratory quotient values induced by controlled feeding of the rat. *Amer. J. Physiol.*, 1937, **120**, 458–465.

121. WILBUR, K. M. A method for the measurement of activity of small animals. *Science*, 1936, **84**, 274.

122. YOUNG, W. C., & FISH, W. R. The ovarian hormones and spontaneous running activity in the female rat. *Endocrinology*, 1945, **36**, 181–189.

123. ZIEVE, L. Effects of benzedrine on activity. *Psychol. Rec.*, 1937, **1**, 393–396.

124. ZIEGLER, L. H., & KNUDSON, A. Activity after recovery from rickets. *J. comp. Psychol.*, 1936, **22**, 199–217.

SAMPLING IN THE REVISION OF THE
STANFORD-BINET SCALE

ELI S. MARKS
National Office of Vital Statistics

In another paper (4) the writer attempts to point out the biases which may arise through types of sampling procedure quite common in psychological research. The present analysis is devoted to another effect of sampling methods commonly used in psychology—namely, the substantial increase in sampling error which results when "cluster" methods of sampling are used. It should be noted that this is not a criticism of the cluster type of sampling. Cluster sampling is an extremely valuable device and makes feasible many studies which would otherwise be completely impossible. However, the use of cluster techniques implies substantial modifications in our formulae for sampling error and psychologists are, in general, not familiar with these modifications. Unfortunately, much of the important work in the field appears in sources which are relatively inaccessible to psychologists. Ignoring the effects of cluster sampling on measures of sampling error has undoubtedly resulted in attaching importance to results which are statistically insignificant. In the testing field, failure to allow for cluster sampling has probably caused us to attach a measure of precision to our results considerably in excess of that warranted by sound statistical techniques.

Cluster sampling almost always involves an increase in sampling error as compared with unrestricted random sampling of the same number of cases. It is, of course, possible to obtain a lower sampling error with cluster sampling than with unrestricted random sampling if we make up our clusters for this purpose. However, the main reason for the use of cluster sampling is to permit the sampling of previously existing groups (the clusters) and, in most cases, the use of a previously existing grouping of the population involves a positive intraclass correlation of the variable studied, i.e., our existing groups are almost always more homogeneous internally than groups of the same size made up by random selection of individuals from the population. It is the existence of positive intraclass correlation which cuts down the amount of independent information available from a cluster sample of a specified size and occasions the substantial increase in sampling error usually associated with this sampling method. The present analysis is designed to emphasize the substantial increase in sampling error which results from relatively small intraclass correlations. While this phenomenon is quite familiar to sampling statisticians, psychologists are rather generally

413

unaware of the marked disturbances of sampling error calculations and tests of significance introduced by the use of cluster sampling when a positive intraclass correlation exists.

Although methods resembling cluster sampling are quite common in psychological research, very few psychological studies have used sampling designs which permit us to determine the standard error of the mean or of other sample statistics. As a matter of fact, it is difficult to find a study where analysis of the sampling error formulae used is not complicated by the presence of a non-measurable design (one in which the sampling probabilities are unknown). Some of the difficulties in the use of non-measurable designs are explored in a paper by McNemar (2) which discusses accidental sampling and purposive sampling as well as such measurable designs as unrestricted random sampling and stratified sampling.

THE SAMPLING PLAN OF THE STANFORD-BINET

The writer has, therefore, not attempted to find a study with a measurable design, but has selected for analysis the sample used in the revision of the Stanford-Binet. This sample has been selected for analysis principally because the widespread use of the revised Stanford-Binet makes the problems involved in its standardization extremely important in spite of the lapse of a decade since the revision was completed. The revision of the Stanford-Binet is also a good example for our purposes because (a) it was an extensive project, involving a relatively large number of subjects and the expenditure of considerable amounts of time, effort and money and (b) the purposes of the sample were explicitly formulated and clearly stated by the authors of the revised Stanford-Binet.

The reader should bear in mind that the present analysis is on an "as if" basis. The Stanford-Binet sampling design does not yield statistics with measurable standard errors and no amount of statistical manipulation can overcome this defect. The cure lies not in statistical formulae but in more careful sampling techniques in future investigations. However, the use of measurable sampling designs in psychological research will almost inevitably mean cluster sampling of some sort since any other approach will be beyond the limited resources usually available to psychologists. Thus, an examination of cluster sampling, even on an "as if" basis, is extremely pertinent to the future of any psychological research which involves statistical techniques.

In my analysis I have relied entirely upon statements and data published in Terman and Merrill (6) and McNemar (3). Since the data

required for this analysis have not been published in full detail, I have been forced to use approximations at several points. Inquiry indicated that more detailed data could not be furnished without considerable expenditure of time and effort. Since the approximations used in this paper are satisfactory for purposes of illustration and since the sampling techniques used in the revision of the Stanford-Binet preclude a completely accurate determination of error even if the detailed data were available, this deficiency is not serious. In nearly every case, the effect of the approximation used has been to understate the sampling error.

In revising the Stanford-Binet, the major objective was to construct scales "so standardized for difficulty as to yield mean I.Q.'s of approximately 100 at all age levels" (Terman, in 3, p. 3). The authors of the revision realized that their success in this objective was dependent upon securing a measure of the distribution of test scores in the general population (or in a satisfactory sample of the population). The sample was restricted to "American born" subjects of the "white race" in the age range from $1\frac{1}{2}$ years to 18 years. Terman notes that "elaborate precautions were taken to make the sampling as representative of the entire population as circumstances permitted" (3, p. 6).

According to Terman and Merrill this was done by selecting "17 different communities in 11 states" (6, p. 12). They note that: "The selection of localities for the second year's testing was based upon certain considerations in regard to sampling which had resulted from a study of the socio-economic level of the first 1500 subjects." These considerations were what the authors viewed as an inadequate representation of the rural group and a difference between the occupational distribution of fathers of the cases tested and the occupational distribution of all employed U. S. males. In the second year's testing, therefore, the authors of the Stanford-Binet revision "took care to include several additional rural communities" (6, p. 14) Neither McNemar nor Terman and Merrill give further details on the method of selecting the communities. It seems evident that selection was not on the basis of random sampling (neither simple random sampling nor random sampling within strata). As a matter of fact the term "community" is not defined clearly enough to permit a rigorous statement of the primary sampling units used. Nevertheless, we can visualize our population as being composed of "communities" (undefined but definable), so that the entire population of the United States can be broken up into a fairly large number (probably over 3000) of communities.

Within each community different procedures were followed for cases in the three age groups—$1\frac{1}{2}$ to $5\frac{1}{2}$ inclusive; 6 to 14 inclusive; and 15 to 18 inclusive. These groups were sampled as follows:

1. *The group aged 6 to 14.* Schools of "average social status" were selected in each community (method of selecting schools not further specified) and within each school all of the children between the ages 6 to 14 who were within one month of a birthday were taken, regardless of grade placement (6, p. 15). This sampling procedure is, then, a subsampling of subclusters with a 100 percent sample take within the subcluster.

2. *The group aged 15 to 18.* Subjects were selected so that "the advanced group would be as nearly as possible continuous with the intermediate, with no break between fourteen and fifteen years. The compulsory school age was taken into account, the general character of the population, and the type of secondary education that was offered. In each community the school census was consulted to determine the amount of elimination after age fourteen. We made certain that some of the twelve-, thirteen-, and fourteen-year-olds who had gone to high school were included, also some of the slow fifteen- and sixteen-year-olds who were still in intermediate school. A few cases who had graduated from high school were included and a few who had dropped out of school without completing high school. These out-of-school groups were sampled by choosing siblings of school children in numbers proportional to the amount of elimination at ages above fourteen." (Sampling in this group was actually a rough type of "quota" sampling.)

3. *The group aged 1½ to 5½.* This group was sampled in much the same manner as the out-of-school cases in the group aged 15 to 18. The authors "chose as far as possible younger sibs of the school groups." Children were secured by use of birth records, school census, school siblings, kindergartens, well baby clinics, day nurseries, nursery schools and "personal report." Use of the various sources differed from community to community. "Great care was exercised in the large population centers to include representative groups; if a school in a suburban district which had been chosen as average on the advice of superintendent and counselors seemed to include too large percentage of higher occupational groups it was offset by a tenement district center." The authors state that "in the smaller communities, from seventy-five to eighty percent of the pre-school child population of appropriate age was examined" (6). In the published tabulations results for children aged 1½ were omitted and further references deal only with the sample of children two years of age or over.

It may be noted that the population sampled is limited to individuals within one month of a birthday (or half-year birthday for children under six). The population is also limited to American-born white persons and, in the age range six to 14, to children attending school. These limitations do not affect the propriety of generalization from a sample to the population so defined. The limitations may affect generalization from the sample to all native-born white persons aged two to 18. This is not, however, of primary concern in this paper. Limitations on generalization resulting from the use of sub-populations are discussed by the writer in another article (4). For our present purposes, it is sufficient to accept the population, as defined.

To summarize, the sampling plan of the Stanford-Binet revision

involved: (a) sampling of "communities" from the aggregate of all United States communities; (b) the subsampling of schools for children aged six to 14 and taking all children (in the population as defined above) in the selected schools; (c) the subsampling of other members of the defined population from the "community" without any intermediate subsampling of schools but with the use of a rough type of "quota" sampling.

BIASES AND VARIANCE IN THE STANFORD-BINET SAMPLING

The above is only an approximate statement since it is extremely hard to formulate exactly the sampling plan used. The method of selection at each stage of sampling has not been specified above. It seems likely, however, that the sampling error of the plan used is greater than the error which would be involved in random sampling of "communities" with equal probability of selection and no subsampling (i.e., a plan which would take all persons in the communities selected).

It is obvious that a sampling plan not involving subsampling will have a lower sampling error for the same number of clusters sampled than a plan which did involve subsampling. The assumption that the community sampling actually used involved a larger sampling error than random selection is not as clear cut. Actually the sampling resembled "purposive sampling" or "quota sampling" but it does not appear to conform even to the rather loose requirements of these two techniques.

In discussing purposive sampling Neyman (5) developed certain hypotheses which, if satisfied, would make the estimate commonly used in this method the "best linear estimate" (i.e., an unbiased linear estimate with variance less than that of any other linear estimate). Neyman notes that:

If these hypotheses are not satisfied, which I think is a rather general case, we are not able to appreciate the accuracy of the results obtained. Thus this is not what I should call a representative method. Of course it may sometimes give perfect results, but these will be due rather to the uncontrollable intuition of the investigator and good luck than to the method itself.

While the Stanford-Binet revision did not involve purposive sampling, Neyman's remarks are applicable to the sampling plan. Furthermore, there is internal evidence in the results of the Stanford-Binet revision which indicates that, in spite of the purposive attempt to secure a "representative" sample, the Stanford-Binet revision sample actually produced a larger sampling error than would have resulted from random sampling of clusters.

Table 3 below gives the number of cases from each of the communities included in the Stanford-Binet sample. It should be noted that 37 percent of the "urban" cases were drawn from San Francisco and 56 percent of the "suburban" cases came from two California communities. This means that 975 cases or 34 percent of the total sample were from California. In addition a disproportionately large number of the "rural" cases (41 percent) came from one community in Vermont. It would, of course, be possible to obtain clusterings in two states as marked as those shown by a random sampling of communities, but the probability of such an outcome is extremely small. It is almost certain that a random (or stratified random) sampling of communities would have given a better geographic distribution (and undoubtedly a lower sampling error) than was actually obtained. This fact is also pointed out by McNemar (3) who expresses "skepticism concerning the representativeness of these communities."

It should also be remembered that the authors of the Stanford-Binet revision felt very definitely that their results contained a substantial bias. As noted above, the primary objective of the revision was to obtain a scale giving average I.Q.'s of 100 for each chronological age group. Terman and Merrill (6, p. 23) note that the mean I.Q.'s for their age groups run "slightly above 100" and state that this "is the result of intentional adjustment to allow for the somewhat inadequate sampling of subjects in the lower occupational classes." McNemar (3, p. 20) states:

The fact that the means in Tables 1 and 2 are above 100 should not lead the reader to the erroneous conclusion that the average I.Q. for the population now exceeds 100. The excess here observed is in the proper direction to allow for known bias in our age samplings. When an adjustment is made for bias in occupational status, the age means approach nearer 100, and a further adjustment for inadequate rural representation would tend to bring the values still closer to 100.

Table 6 on p. 36 of Terman and Merrill (6) gives average I.Q.'s for each age group "adjusted for 1930 Census frequencies of Occupational groupings." These averages still show substantial bias, all means except those for ages 4 and 5½ being over 100 and seven age groups having average I.Q.'s over 103. The effect of rural-urban biasing influences is not presented.

Since the method of correcting for bias is not stated, the effect of these corrections on the mean square errors of the sample results cannot be determined. It is probably not possible to make this determination in any event since the presence or absence of biases in occupational or

rural-urban distributions cannot by themselves tell us whether an I.Q. distribution is biased or unbiased and correcting for rural-urban or occupational biases may have very little effect (or even an unfavorable effect) upon I.Q. biases.

In any event, the original sample means of the Stanford-Binet revision contain substantial biases if the true population means are 100. These are shown by the figures in Table 1.*

TABLE 1

AVERAGE I.Q.'s BY AGE GROUPS FOR THE STANFORD-BINET REVISION SAMPLE

	Age Groups			
	2½–5½	6–13	14–18	All Cases
From L—Mean	106.58	103.22	103.03	104.00
Form M—Mean	106.42	103.96	103.32	104.43
Number of Cases	728	1623	619	2970

In view of the probable biases and the considerations with regard to the sampling method presented above, it is not at all unreasonable to assume that the Stanford-Binet revision sampling involved a larger standard error of the mean than would random selection of communities with equal probability. Even if this is not the case, the subsampling involved should account for an increase in sampling error over a design in which there was no subsampling.

On the basis of the above discussion, the standard error of random selection of communities with equal probability and no subsampling gives us minimum values for the standard errors of the Stanford-Binet sample means. To estimate these errors, we shall assume that the number of cases actually sampled in each community was the total eligible population in that community. (As noted above, assuming that the community population was larger than the number sampled would lead to a larger estimate of the standard error.)

---

* The data are from McNemar (3) Tables 1 and 2. There are minor differences between the results presented by McNemar and those presented by Terman and Merrill (apparently due to inclusion of some subjects in some of the distributions and their omission in other distributions). The differences are minor and do not affect the present analysis.

The data in this and in the two subsequent tables are reproduced with the permission of Houghton Mifflin Co., the publishers of McNemar's *The Revision of the Stanford-Binet Scale.*

The Standard Error for Cluster Sampling

The standard error for the type of sampling described (i.e. "cluster sampling") is given by:

$$\sigma_{\bar{x}'}^2 = \frac{M-m}{(M-1)m} \frac{\sum_i^M N_i^2(\bar{x}_i - \bar{x})^2}{M\bar{N}^2}. \qquad [1]$$

Or, when we estimate $\sigma_{\bar{x}'}$ from the sample, the estimated standard error is given by:

$$s_{\bar{x}'}^2 = \frac{M-m}{Mm} \frac{\sum_i^m N_i^2(\bar{x}_i - \bar{x}')^2}{(m-1)(\bar{N}')^2} \qquad [2]$$

where M = the total number of clusters (communities) in the population

m = the number of communities sampled

N_i = the number of individuals (eligible for the population) in the i-th cluster

$\bar{x}_i$ = the mean I.Q. for the N_i individuals in the i-th cluster

$$\bar{x} = \frac{\sum_i^M N_i \bar{x}_i}{\sum_i^M N_i} = \text{the mean I.Q. of the population. (The aim of the sample is to estimate } \bar{x}.)$$

$$\bar{N} = \frac{\sum_i^M N_i}{M} = \text{the average number of individuals per cluster in the population.}$$

$$\bar{x}' = \frac{\sum_i^m N_i \bar{x}_i}{\sum_i^m N_i} = \text{the mean I.Q. of the sample. (We are using this as our estimate of } \bar{x} \text{ and } s_{\bar{x}'} \text{ is the estimated standard error of this sample mean.)}$$

$$\bar{N}' = \frac{\sum_i^m N_i}{m} = \text{the average number of individuals per cluster in the sample.}$$

To determine $s_{\bar{x}'}$ exactly we need to know M, the number of communities (clusters) in the population. While M is not known with any precision, we can be quite certain that it is large and that it is much

larger than m (at least 100 times as great). Consequently, we can, without appreciable error, take $M - m/M$ equal to 1. With this substitution the square of the standard error is approximately equal to:

$$ s_{\bar{x}'}{}^2 = \frac{\sum\limits_i^m N_i{}^2(\bar{x}_i - \bar{x}')^2}{m(m-1)(\overline{N'})^2} = \frac{m}{m-1} \cdot \frac{\sum\limits_i^m N_i{}^2(\bar{x}_i - \bar{x}')^2}{\left(\sum\limits_i^m N_i \right)^2}. \qquad [3] $$

All the data required for Equation [3] can be obtained from the sample. Unfortunately, not all of the sample data are available in published form. Since we shall have to rely on published data, some further approximations (described below) are necessary. The approximations also act to reduce our estimate of the standard error.

McNemar (3) gives, as Table 9, information on the average I.Q.'s for children in "urban," "suburban" and "rural" communities by age groups. This table, plus data for the entire group in the age range 2 to 18, is presented in Table 2. The data for the entire group were calculated from the information given for the three age groups.

TABLE 2

I.Q. DATA FOR URBAN, SUBURBAN AND RURAL CHILDREN*

	Urban	Suburban	Rural	Total
2–5½ Year-Olds				
Number	354	158	144	656
Mean	106.3	105.0	100.6	104.7
S.D.	15.7	16.1	15.4	15.9
6–14 Year-Olds				
Number	864	537	422	1823
Mean	105.8	104.5	95.4	103.0
S.D.	14.7	16.8	15.5	16.1
15–18 Year-Olds				
Number	204	112	103	419
Mean	107.9	106.9	95.7	104.6
S.D.	16.5	15.7	15.9	16.9
All Ages (2–18)				
Number	1422	807	669	2898
Mean	106.2	104.9	96.6	103.6
S.D.	15.2	16.5	15.7	16.2

* Denver 2- to 5½-year-olds are excluded.

To determine $s_{\bar{x}'}$ we shall take for our values of $\bar{x}_i$ (the mean I.Q. in each community): (a) the average I.Q. for urban children for each of

the communities classified as "urban by McNemar; (b) the average I.Q. for suburban children for each of the communities classified as "suburban" and (c) the average I.Q. for rural children for each of the communities classified as "rural." This approximation ignores all variations between communities within the urban, suburban and rural groups of communities. As a result the values of $s_{2'}$ which we obtain should be equal to or less than the values which would be obtained if we knew the means of each of the sampled communities.*

There are some uncertainties in the published data concerning the values of m and N_i. As noted above, Terman and Merrill (6) state that 17 communities were sampled in 11 states. This would give $m = 17$. However on pp. 36–37, McNemar (3) lists the communities sampled and the number of subjects in each community. McNemar lists 7 urban communities. He also lists 3 suburban communities and states that, in the suburban group, there were "four small communities just out of Kansas City in Johnson County Kansas, with 199 cases drawn from Westwood View, Hickory Grove, Roseland, and Shawnee Mission schools." For the rural communities, McNemar states:

> The samplings from rural communities include 85 from Mount Washington School, Bullitt County, and Liberty School, Oldham County, Kentucky. A total of 152 were drawn from the following districts of Indiana: Prather School, Charlestown schools and Morgan Township School in Harrison County and Galena School in Floyd County. A farming region at Bloomington, Minnesota, supplied 92 cases; the farming and small village community of Randolph, Vermont, provided 275; and 65 subjects were secured in the vicinity of Atlee, Virginia. We have already expressed some skepticism concerning the representativeness of these communities.

From this statement, it is difficult to determine the exact number of "rural communities" involved. At a minimum, there appear to be 8 (assuming that schools in different counties represent different communities). If we also consider the four schools in the "suburban" part of Johnson County, Kansas, to be one community, McNemar's listing gives a count of 19 communities vs. Terman and Merrill's 17. The difference appears to be one in the definition of community. In terms of independently selected areas, Terman and Merrill's "17 communities" is probably more nearly correct. However, the data in Table 2 are based on McNemar's classification. It appears desirable to adopt a compromise, counting as communities the cities and towns listed by McNemar

---

* This statement cannot be made absolutely since, under certain circumstances, it may be incorrect. However, it is a fairly safe statement since the circumstances which would give a higher standard error through substituting group averages for individual averages are extremely unusual.

plus any schools in separate counties. This is the same basis we used in getting the count of 19 communities mentioned above. Since the count of independently sampled communities is probably Terman and Merrill's figure of 17, this handling of the problem operates in the same direction as the other approximations previously made.

The difficulty in determining N_i occurs in the cases where McNemar gives one figure for the number sampled in two different counties (e.g.,

TABLE 3

NUMBER OF CASES SAMPLED IN EACH COMMUNITY AND
ESTIMATED AGE DISTRIBUTION

Communities	2–5½ Year-Olds	6–14 Year-Olds	15–18 Year-Olds	All Ages (2–18)
Urban				
1. Denver, Col.	28	67	16	111
2. Minneapolis, Minn.	46	111	26	183
3. New York, N. Y.	12	29	7	48
4. Reno, Nev.	28	68	16	112
5. Richmond, Va.	46	114	27	187
6. San Antonio, Texas	63	155	36	254
7. San Francisco, Calif.	131	320	76	527
Suburban				
8. White Plains, N. Y.	31	107	22	160
9. Redwood City, Calif.	26	89	19	134
10. Los Gatos, Calif.	62	209	43	314
11. Johnson County, Kan.	39	132	28	199
Rural				
12. Bullit County, Ky.	9	27	7	43
13. Oldham County, Ky.	9	27	6	42
14. Clark County, Ind.	11	32	8	51
15. Harrison County, Ind.	11	32	8	51
16. Floyd County, Ind.	11	31	8	50
17. Bloomington, Minn.	20	58	14	92
18. Randolph, Vt.	59	174	42	275
19. Atlee, Va.	14	41	10	65

85 cases from Bullitt County, Kentucky and Oldham County, Kentucky). These cases can be handled by distributing the cases equally among the counties involved. This adjustment also operates to reduce the estimated standard error. A further approximation is necessary to get standard errors for the means of each of the three age groups in Table 2. McNemar gives only the total number of cases in each community and does not give the distribution of these cases among the age groups. To estimate the standard errors for the separate age groups, the number

of cases for each of the communities was distributed by age proportionately to the age distribution in the class (urban, suburban or rural) in which the community falls. The number of cases in each community shown by McNemar and the calculated distribution of these cases by age groups is shown in Table 3. This adjustment affects only the estimates of the standard errors of the age group averages and not the standard error for the entire group aged 2 to 18.

COMPARISON OF CLUSTER SAMPLING ERROR WITH UNRESTRICTED RANDOM SAMPLING ERROR

With all the adjustments reducing the standard error which have been made, it may seem surprising that we have any error left. However, a fairly substantial amount of sampling error remains. Table 4 shows the standard errors of the mean I.Q. calculated as described above (using Equation 3) compared with the standard error obtained by the formula usually used in psychological research studies, i.e.:

$$\sigma_{\bar{x}'}{}^2 = \frac{\sigma^2}{N'} \tag{4}$$

where

$$\sigma^2 = \frac{\sum\limits_{i}^{m} \sum\limits_{j}^{N_i} (x_{ij} - \bar{x}')^2}{N'} \tag{5}$$

and

$$N' = \sum\limits_{i}^{m} N_i. \tag{6}$$

In Equations [4], [5] and [6], x_{ij} stands for the value (I.Q.) of the jth individual in the ith cluster (community) and the other symbols have the meanings previously defined. Equation [4] represents the standard error of the mean of a sample drawn by unrestricted random sampling from an infinite population (i.e. a sample drawn so that the probability of drawing any observation in the population is equal to and independent of the probability of drawing each of the other observations).

It will be seen from Table 4 that the absolute values of the standard errors calculated by Equation [3] are not large. There is a sampling error of only 1 per cent in the average I.Q. for the entire group of 2,898 cases. However, a very substantial difference exists between the stand-

ard error by Equation [3] and the standard error by Equation [4]. If we apply Equation [4] to determine the standard error of the mean of a cluster sample, it is obvious that we shall be very far from the correct value (in this case we would get an error which is less than one-third of the correct figure).

This fact is extremely important in applying tests of significance to differences of sample means. For example, suppose we took a sample

TABLE 4

ESTIMATED STANDARD ERRORS OF THE MEAN I.Q.'S FOR CLUSTER SAMPLING AND UNRESTRICTED RANDOM SAMPLING

	Standard Errors		Ratio of S.E. of Cluster
Age Groups	*Cluster Sampling*	*Unrestricted Random Sampling*	*Sampling to S.E. of Random Sampling*
2–5½ years	.60	.62	.97
6–14 years	1.09	.38	2.89
15–18 years	1.35	.82	1.63
All Ages (2–18)	1.01	.30	3.36

of 900 children aged 2–18 (by a method which was actually random) from some city or other population subgroup. Assume that this sample gives us an average I.Q. of 105.7 on the revised Stanford-Binet and our sample has a standard deviation of 18, so that the standard error of the mean (using, quite properly, Equation [4]) is .60. Our group has a mean 2.1 points above the average of 103.6 for the Stanford-Binet revision sample shown in Table 2. We want to know whether this difference is significant. If we assume unrestricted random sampling of the Stanford-Binet revision sample, we would use .30 (see Table 4) as the standard error of the revision sample mean. This would give us .67 as the standard error of the difference of 2.1 and our difference would be 3.1 times its standard error. We would undoubtedly consider this a significant difference. Actually, the standard error of the mean of the revision sample is at least 1.01, which makes the standard error of the difference 1.17. The difference is actually only 1.8 times its standard error and can hardly be considered significant.

The sample used for the Stanford-Binet revision is not an extreme case of the error which can be made by applying formulae based on unrestricted random sampling to data obtained by cluster sampling. The sampling for the Stanford-Binet revision did involve testing individuals from several communities and the standard error for cluster sampling is

only 3 times the error for random sampling of the same number of individuals. Many studies use data from one or two groups (e.g.) elementary psychology classes in two neighboring colleges) to draw conclusions about the whole population (all college students or even all human beings). In this case the standard error obtained from Equation [3] may be 50 to 100 times greater than that obtained from Equation [4]. Use of the "correct" formula ("correct" if we have used a random process for drawing our groups) will make supposedly significant differences vanish more rapidly than a quart of ice cream at a children's party.

INTRACLASS CORRELATION

The reason for the difference between the standard error for unrestricted random sampling and that for cluster sampling is to be found in the fact that individuals are not sampled independently in cluster sampling. If we consider samples of equal size from the same population, the standard error of the mean in unrestricted random sampling is multiplied by approximately $(1 + \overline{N}\rho)$ when we use cluster sampling. Here $\overline{N}$ is the average size of our clusters and ρ is the intraclass correlation (a measure of the extent to which individuals within a cluster resemble, or are "correlated" with, each other). The intraclass correlation usually ranges from 0 to $+1$ (although it can be negative). It can be seen that even very small values of the intraclass correlation (say, .01) can have a very substantial effect on the standard error of a mean in cluster sampling if the clusters are moderately large ($\overline{N} = 100$ or more). As a matter of fact, the estimated intraclass correlation for the entire sample (all individuals aged 2 to 18) used in the Stanford-Binet revision is only .08. A recent paper by Walsh (7) gives some of the probability considerations involved in tests of significance when intraclass correlation is present.

There is one feature of Table 4 which may arouse some interest. This is the fact that the estimated standard error (using Equation [3]) of the mean I.Q. is larger for the group aged 6–14 years than for the group aged 2–5½ years. This is, of course, contrary to what we would expect from the description given of the sampling process. To some extent this peculiarity results from our ignoring subsampling in calculating the standard errors. Consideration of subsampling variation would probably increase the standard errors somewhat and would probably increase the standard error more for the group aged 2–5½ years than for the group aged 6–14 years (since there are fewer of the younger children). As a matter of fact, inclusion of subsampling variation might

double the standard error for the mean I.Q. of the group aged 2–5½ but would probably not increase the standard error of the group aged 6–14 more than 10 per cent

Actually Table 4 shows a lower standard error of cluster sampling for the group aged 2–5½ years than for the group aged 6–14 years because there is less variation among the average I.Q.'s of the urban, suburban and rural children for the younger group. This fact may be due to some basic relation between I.Q. variability and age. For example, McNemar (3) gives a table for adjusting I.Q.'s for differing standard deviation of the I.Q. at various ages. He bases this table on the differences actually found in the sample.

Another explanation of the differences in variability between age groups is to be found in the selective nature of the sampling for the Stanford-Binet revision. Selective sampling seems to have been particularly important in the pre-school group. In another article (4), the present writer points out some effects of selective sampling on group means and also notes that selective sampling will usually affect the standard deviation also. It would be very unwise to hypothesize about the difference between age groups shown in Table 4 unless we had much more information about the sampling probabilities.

This article has used the Stanford-Binet only as an illustration of the dangers of ignoring the intraclass correlation when we are dealing with a cluster type of sampling. In view of the qualifications placed on our analysis, it is not possible to draw any conclusions about the reliability or unreliability of the revised Stanford-Binet as a measuring instrument. There may be good reasons for supposing that the precision of the revised Stanford-Binet is considerably less than many of its users assume. From the sampling standpoint, the sample design used in the revision of the Stanford-Binet was a non-measurable design and there is no way of telling how "bad" or "good" the results were. It has been suggested that the sampling errors shown in Table 4 are probably minimum figures. However, the results do offer a possibility of improving the sampling design in the event that the Stanford-Binet is revised again in the future. An error of 1 I.Q. point in the average I.Q. may not be too serious. If this is the case, the biases in the Stanford-Binet average I.Q.'s could probably be removed by using a sound sample design without any need for an increase in either the number of communities covered or the number of subjects tested. If greater accuracy than a mean correct within 1 per cent is considered necessary or desirable, this could probably be achieved by increasing the number of communities sampled without increasing to any great extent the total

number of subjects tested. As a matter of fact, increasing the number of subjects tested would probably add very little to the accuracy of the final results (at least for the age group 6–14 years). The standard error of a mean in cluster sampling decreases (approximately) in proportion to the square root of the number of clusters sampled. The standard error shown in Table 4 for unrestricted random sampling is .3 of an I.Q. point. The standard error for cluster sampling is 3.36 times this value. Therefore, to get a standard error of .3 using cluster sampling, we would need about 11 times as many communities or about 200 communities. This estimate of the number of communities required is, of necessity, unreliable, since we were forced to estimate our standard errors from a sampling plan which is actually non-measurable, and measuring the non-measurable puts an obvious strain on epistemology.

In designing a sampling plan for a revision of the Stanford-Binet recent developments in sampling theory and practice can be used to increase accuracy without increase in survey costs. The reader's attention is directed particularly to the work of Hansen and Hurwitz (1) in this field. Using the techniques developed by Hansen and Hurwitz, persons revising the Stanford-Binet would probably get satisfactory precision from a well-designed sample of 25 to 100 communities with only a very small increase (if any) in the total number of cases tested.

SUMMARY

This article stresses the dangers of ignoring the intraclass correlation of the population when "cluster" methods of sampling are used. The increase in sampling error resulting from cluster sampling is demonstrated by an analysis of the results of the sample used in the revision of the Stanford-Binet. This sample actually yields "non-measurable" results, i.e. results which do not permit determination of the standard error of the sample mean. However, it is estimated that the standard error of the average I.Q. of this sample is at least 3 times the error which would be calculated by the use of the formula for unrestricted random sampling from an infinite population. The latter formula is the one familiar to psychologists and the one usually used by them regardless of the type of sampling involved. The illustration indicates that very substantial errors may result from this practice and that many results will be considered statistically significant where such a conclusion is entirely unwarranted.

BIBLIOGRAPHY

1. HANSEN, M. H., & HURWITZ, W. N. On the theory of sampling from finite populations. *Ann. math. Statist.*, 1943, **14**, 333–362.

2. McNEMAR, QUINN. Sampling in psychological research. *Psychol. Bull.*, 1940, **37**, 331–365.

3. McNEMAR, QUINN. *The revision of the Stanford-Binet Scale.* Boston: Houghton Mifflin, 1942.

4. MARKS, ELI S. Selective sampling in psychological research. *Psychol. Bull.*, 1947, **44**, 267–275.

5. NEYMAN, JERZY. On the two different aspects of the representative method: the method of stratified sampling and the method of purposive selection. *J. roy. statist. Soc.*, n. s., 1934, **97**, 558–606.

6. TERMAN, L. M., & MERRILL, MAUD A. *Measuring intelligence.* Boston: Houghton Mifflin, 1937.

7. WALSH, J. E. Concerning the effect of intraclass correlation on certain significance tests. *Ann. math. Statist.*, 1947, **18**, 88–96.

APPENDIX

Although the formula for the standard error of the mean for cluster sampling is not new, psychologists are generally unfamiliar with it. The derivation of this formula is, therefore, presented below. The z transformation will be found useful in deriving standard errors for more complicated designs (e.g. designs using stratification, subsampling, differential sampling probabilities, etc.).

Equation [2] gives the mean square error (square of the standard error) of the mean of a cluster sample as:

$$s_{\bar{x}'}^2 = \frac{M - m}{Mm} \frac{\sum_i^m N_i^2 (\bar{x}_i - \bar{x}')^2}{(m - 1)(\overline{N}')^2}.$$

The mean of the sample is:

$$\bar{x}' = \frac{\sum_i^m N_i \bar{x}_i}{\sum_i^m N_i}.$$

M, m, N_i, $\bar{x}_i$, $\bar{x}'$, $\overline{N}'$, $\bar{x}$ and $\overline{N}$ are defined on p. 420 It is also convenient to define:

$$x_{ij} = \text{value for } j\text{th individual in } i\text{th cluster}$$

$$x_i = \sum_i^{N_i} x_{ij} = \text{sum of the values for all individuals in } i\text{th cluster.}$$

From their definitions, it can be seen that:

$$\bar{x}_i = \frac{x_i}{N_i}$$

and, therefore

$$\bar{x}' = \frac{\sum\limits_{i}^{m} x_i}{\sum\limits_{i}^{m} N_i}.$$

$\bar{x}'$ can be treated as a ratio of two linear functions of the sample observations, namely:

$$f(x) = \frac{M}{m} \sum_{i}^{m} x_i = \frac{M}{m} \sum_{i}^{m} \sum_{j}^{N_i} x_{ij}$$

$$f(N) = \frac{M}{m} \sum_{i}^{m} N_i = \frac{M}{m} \sum_{i}^{m} \sum_{j}^{N_i} 1.$$

In deriving $s_{\bar{x}'}{}^2$, it will be useful to prove the following theorem:

Theorem: If we have a sample estimate:

$$r' = \frac{f(x)}{f(y)}$$

where $f(x)$ and $f(y)$ are linear functions of the sample observations x_h and $y_h (h = 1, 2 \cdots n)$ and if:

$$x = Ef(x), \qquad y = Ef(y), \qquad z_h = \frac{x_h}{x} - \frac{y_h}{y}$$

then:

$$\sigma_{r'}{}^2 = \left(\frac{x}{y} - r\right)^2 + \left(\frac{x}{y}\right)^2 \sigma_{f(z)}{}^2$$

where $\sigma_{r'}{}^2$ is the mean square error of r' and r is the population parameter of which r' is an estimate.

Proof: When we have a sample estimate $r' = f(x)/f(y)$, the mean square error of r' can be found by: (a) expanding r' as a Taylor series around x and y (the expected values of $f(x)$ and $f(y)$); (b) subtracting r (the true value of r' for the entire population) from both sides of the equation; (c) squaring both sides of the resulting equation and (d) taking the expected value of the resultant. If we ignore, in our Taylor series, terms involving partial derivatives higher than the first, the result of this operation will be:

$$\sigma_{r'}{}^2 = E(r' - r)^2$$

$$= \left(\frac{x}{y} - r\right)^2 + \left(\frac{x}{y}\right)^2 \left(\frac{\sigma_{f(z)}{}^2}{x^2} + \frac{\sigma_{f(y)}{}^2}{y^2} - 2\frac{\sigma_{f(z)f(y)}}{xy}\right). \quad [7]$$

If we let

$$z_h = \frac{x_h}{x} - \frac{y_h}{y}$$

and if f is a linear function, then:

$$f(z) = \frac{f(x)}{x} - \frac{f(y)}{y} \qquad [8]$$

and

$$Ef(z) = \frac{Ef(x)}{x} - \frac{Ef(y)}{y} = \frac{x}{x} - \frac{y}{y} = 0 \qquad [9]$$

$$\sigma_{f(z)}{}^2 = \frac{\sigma_{f(z)}{}^2}{x^2} + \frac{\sigma_{f(y)}{}^2}{y^2} - \frac{2\sigma_{f(x)f(y)}}{xy}. \qquad [10]$$

Therefore:

$$\sigma_{r'}{}^2 = \left(\frac{x}{y} - r\right)^2 + \left(\frac{x}{y}\right)^2 \sigma_{f(z)}{}^2. \qquad [11]$$

The above theorem can be applied to derive the mean square error of $\bar{x}'$, as follows:

$$\bar{x}' = \frac{f(x)}{f(N)}$$

where

$$f(x) = \frac{M}{m}\sum_i^m x_i \quad \text{or} \quad f(x) = \frac{M}{m}\sum_i^m \sum_j^{N_i} x_{ij}$$

$$f(N) = \frac{M}{m}\sum_i^m N_i \qquad f(N) = \frac{M}{m}\sum_i^m \sum_j^{N_i} N_{ij} \quad \text{where } N_{ij} \equiv 1$$

and

$$Ef(x) = \sum_i^M x_i = x$$

$$Ef(N) = \sum_i^M N_i = N.$$

Let

$$z_i = \frac{x_i}{x} - \frac{N_i}{N} \quad \text{or*} \quad z_{ij} = \frac{x_{ij}}{x} - \frac{1}{N}$$

$$f(z) = \frac{M}{m} \sum_i^m z_i = \frac{\dfrac{M}{m} \sum_i^m x_i}{x} - \frac{\dfrac{M}{m} \sum_i^m N_i}{N} .$$

By Equation [11]:

$$\sigma_{z'}{}^2 = \left(\frac{x}{N} - \bar{z}\right)^2 + \left(\frac{x}{N}\right)^2 \sigma_{f(z)}{}^2 \tag{12}$$

where $\bar{z}$ is the population parameter estimated by $\bar{z}'$ and is:

$$\bar{z} = \frac{\displaystyle\sum_i^M x_i}{\displaystyle\sum_i^M N_i} = \frac{x}{N} .$$

Therefore:

$$\sigma_{z'}{}^2 = \left(\frac{x}{N}\right)^2 \sigma_{f(z)}{}^2. \tag{13}$$

Since $f(z)$ is M/m times a sum of the sample values z_i:

$$\sigma_{f(z)}{}^2 = \frac{M^2(M-m)}{(M-1)m} \frac{\displaystyle\sum_i^M (z_i - \bar{z})^2}{M} \tag{14}$$

where

$$\bar{z} = \frac{\displaystyle\sum_i^M z_i}{M} = Ef(z) = 0. \tag{15}$$

An unbiased estimate of $\sigma_{f(z)}{}^2$ from the sample is:

$$s_{f(z)}{}^2 = \frac{M^2(M-m)}{Mm} \frac{\displaystyle\sum_i^m (z_i - \bar{z}')^2}{m-1} \tag{16}$$

* The result is the same whether the z transformation is applied to the cluster totals or the individual observations.

where

$$\bar{z}' = \frac{\sum\limits_{i}^{m} z_i}{m} = \frac{\sum\limits_{i}^{m} x_i}{mX} - \frac{\sum\limits_{i}^{m} N_i}{mN}.$$

From Equations [13], [14], and [16] we have:

$$\sigma_{\bar{z}'}^2 = \left(\frac{x}{N}\right)^2 \frac{M^2(M-m)}{(M-1)m} \frac{\sum\limits_{i}^{M}(z_i - \bar{z})^2}{M} \tag{17}$$

and

$$s_{\bar{z}'}^2 = \left(\frac{x}{N}\right)^2 \frac{M^2(M-m)}{Mm} \frac{\sum\limits_{i}^{m}(z_i - \bar{z}')^2}{m-1}. \tag{18}$$

In Equation [17] we substitute the values:

$$z_i = \frac{x_i}{x} - \frac{N_i}{N}, \qquad \bar{z} = 0, \qquad \overline{N} = \frac{\sum\limits_{i}^{M} N_i}{M}$$

and get:

$$\sigma_{\bar{z}'}^2 = \frac{(M-m)}{(M-1)m} \frac{\sum\limits_{i}^{M}\left(x_i - N_i \frac{x}{N}\right)^2}{M\overline{N}^2} \tag{19}$$

or

$$\sigma_{\bar{z}'}^2 = \frac{M-m}{(M-1)m} \frac{\sum\limits_{i}^{M} N_i^2(\bar{x}_i - \bar{x})^2}{M\overline{N}^2}. \tag{20}$$

We make the same substitutions in Equation [18] and also substitute for $\bar{z}$ and $\overline{N}$ the sample estimates:

$$\bar{x}' = \frac{f(x)}{f(N)} = \frac{\sum\limits_{i}^{m} x_i}{\sum\limits_{i}^{m} N_i}$$

and

$$\overline{N}' = \frac{f(N)}{M} = \frac{\sum_{i}^{m} N_i}{m}.$$

This gives:

$$s_{\bar{x}'}{}^2 = \frac{M - m}{Mm} \frac{\sum_{i}^{m} N_i{}^2(\bar{x}_i - \bar{x}')^2}{(m - 1)(\overline{N}')^2} \qquad [21]$$

or

$$s_{\bar{x}'}{}^2 = \frac{M - m}{M} \frac{m}{m - 1} \frac{\sum_{i}^{m} N_i{}^2(\bar{x}_i - \bar{x}')^2}{\left(\sum_{i}^{m} N_i\right)^2}. \qquad [22]$$

In some cases, cluster sampling may introduce a substantial bias into the sample standard deviation (when the sample S.D. is used as an estimate of the population S.D.). This bias will be practically eliminated by use of the estimate:

$$s_s{}^2 = \sigma_s{}^2 + s_{\bar{x}'}{}^2 \qquad [23]$$

where σ_s is the sample S.D. and s_s is an estimate of the population S.D.

Equation [23] can also be used for estimating the population S.D. from a sample with unrestricted random sampling.

ILLUMINATION STANDARDS FOR EFFECTIVE
AND EASY SEEING

MILES A. TINKER
University of Minnesota

The problem of artificial illumination is of primary importance in all inside working environments. To maintain healthful and efficient functioning of the eyes, it is necessary to provide adequate lighting. Unquestionably, proper illumination contributes much to comfort and efficiency in activities of daily life. Working under faulty illumination frequently results in eyestrain which tends to be accompanied by reflex functional disturbances of other organs.

During recent years a "lighting consciousness" has been forced upon a large portion of the population, particularly upon those who do considerable visual work under artificial light and upon those who must decide upon the illumination requirements of schools, offices, factories and other situations where visual work is to be performed. Although interest in lighting has been stimulated by popular articles, advertisements, and "educational pamphlets"—as well as by reports written by educators and medical men—the more fundamental information has appeared as experimental reports in scientific publications. The result of exposure to this material is a keen interest in illumination and a sincere desire on the part of the public for sound information concerning hygienic lighting. The natural tendency is to consult pamphlets on recommended practice when lighting specifications are needed for a particular situation. Frequently, the applied psychologist will be called upon to furnish advice on proper illumination. In many instances he will be asked to evaluate the materials presented in the recommended practices. Consequently, the applied psychologist should be informed concerning the adequacy of the data from which the lighting specifications in the recommendations are derived.

The first code on lighting was issued by the Illuminating Engineering Society in 1915. In the more recent publications, the codes are known as Recommended Practice of Home Lighting, of Office Lighting, etc. These pamphlets have been prepared by the Illuminating Engineering Society either alone or jointly with the American Institute of Architects, usually under the rules of procedure of the American Standards Association. Although the American Psychological Association has been in existence for over 50 years, and even though applied psychologists have been interested in the field and have been making experimental contributions to the hygiene of vision for over 40 years, neither psy-

435

chology nor psychologists are represented in the group specifying recommended practices. Furthermore, a large body of psychological literature has been ignored, either because the illuminating engineers were not familiar with it or because they chose not to use it. The result has been an emphasis upon the engineering aspects of lighting with inadequate attention to certain psychological factors. More recently there has been some attempt to consider more of the psychological factors. Perhaps because engineers lack a psychological background, interpretations are frequently erroneous. Probably the most satisfactory approach to hygienic lighting could be achieved by coordinating the contributions of engineers, physiologists, and psychologists.

Recent editions of recommended practices reveal an increased emphasis upon control of direct and reflected glare, brightness contrast, and the diffusion or distribution of light. The tendency to specify relatively very intense light for many visual tasks is prominent. The purpose of this paper is to present a critical examination of the specifications in the more recent editions of recommended practices and to scrutinize some of the data from which the recommendations were derived.

SPECTRAL QUALITY OF LIGHT

In general, spectral quality of light receives adequate treatment in recommended practices (35, 36, 37, 38). It is stated that with equal foot candles of illumination, variations in color quality of light found in common illuminants have little or no effect upon the visual discrimination involved. When color is to be discriminated, it should be viewed under as close an approximation of daylight as possible. Luckiesh (10) has a valuable discussion of light and color.

QUALITY OF LIGHTING

Recommendations (35, 36, 37, 38) concerning control of glare, diffusion, direction and distribution of light, light reflection value, and effects of finishes on ceilings and wall are ordinarily quite satisfactory. Visual discrimination is improved by moving the glare source away from the line of vision and by reducing the brightness of the light source and the amount of light emitted by the light source toward the eye. Brightness of luminaires should be low in value. High brightness contrasts within the field of vision should be avoided whether on the work surface or in other parts of the visual field. Proper diffusion of light helps to eliminate undesirable shadows. Purely local lighting, therefore, is unsatisfactory. Since the reflection factors of objects in the visual environment play an important role in illumination, the finish of ceilings, walls,

floors and furnishings is important. These surfaces should provide reflecting surfaces to help spread the light about the room. Furthermore, they should be such that undesirable brightness contrast does not occur within the field of vision. Shiny or glossy finishes should be avoided to prevent specular glare.

In the recommended practices, informative discussions on classification of lighting systems are usually included. Also illustrations of fixtures and installations are sometimes given. Some attention is given to daylight illumination and the need of coordinating artificial with daylight lighting.

INTENSITY OF ILLUMINATION

Intensity of illumination receives by far the greatest emphasis in specifications. With each revision of a lighting code prepared by illuminating engineers, the foot candle recommendations for a given situation rise. One may well question whether this trend has a scientific basis, or whether the consumer has been educated to accept the higher intensities. In 1934, Luckiesh and Moss (11) presented general recommendations which they considered to be very conservative. These are repeated with slight changes in Luckiesh's 1944 book (10). He adds that these are inadequate in many cases where hundreds and even thousands of foot candles of light are desirable. Examination of the recommended practices of lighting reveals that, for the most part, they are based upon researches done and interpretations made by Luckiesh and his co-workers, or upon researches inspired by them. Let us turn first, therefore, to these reports.

In *Light, Vision and Seeing*, Luckiesh (10), and in the *New Science of Seeing*, Luckiesh and Moss (11), make the following foot candle recommendations for common tasks of the work-world:

1. 100 foot candles or more are specified for severe and prolonged visual work. Examples include fine needle work, pen work, engraving and assembly, and discrimination of fine details involving low contrast.

2. 50 to 100 foot candles should be used for proof-reading, difficult reading, watch repairing, and average sewing.

3. 20 to 50 foot candles are listed for such visual tasks as clerical work, ordinary reading and average sewing on light goods.

4. 10 to 20 foot candles are proposed for ordinary reading and sewing on light goods when the task is not prolonged.

5. 5 to 10 foot candles are needed for visual work which is more or less interrupted or casual.

6. 1 to 5 foot candles are sufficient for perceiving large objects.

Luckiesh (10) states that these are minimum foot candle recommendations and that he considers them to be very conservative from the

viewpoint of ease of seeing. Furthermore these foot candles, according to Luckiesh and Moss (11), are far below the intensities of illumination which new knowledge indicates to be ideal.

These recommendations are derived from various sets of data which will be discussed in turn.

Preferences for light intensity. Luckiesh and Moss (11) cite data on preferences for light intensities to support their contentions that high intensities are necessary for adequate seeing. The mean choice was about 100 foot candles but the median was 50 foot candles when up to 1000 foot candles were available. Tinker's analysis (22) of light preference studies indicated that visual adaptation plays an important role in determining the preferences. In an experimental check, Tinker (26) found that when readers were adapted to 8 foot candles, the median choice for comfortable reading was about 12 foot candles. But when adapted to 52 foot candles, the median choice was 52 foot candles. It is obvious that the intensity of illumination to which the reader is adapted plays a dominant role in his illumination preference. The conclusion is, therefore, that preference for illumination intensity is not a satisfactory method for determining the intensity of light needed for efficient visual work.

Visual acuity. Luckiesh and Moss (11) and Luckiesh (10) list visual acuity as a basic factor in reading (and presumably in other visual work). It is true enough that visual discrimination does depend somewhat upon visual acuity. But is visual acuity an adequate criterion for prescribing appropriate lighting? Luckiesh and Moss (13) admit that in many tasks the criterion of visual acuity is relatively inappropriate, e.g. in tasks involving low contrasts. But they point out that for black test objects on a white background, visual acuity improves up to 100 foot candles. As a matter of fact, Lythgoe (15) has shown that under certain conditions of measurement, visual acuity improves up to and beyond 1000 foot candles. Inspection of the data reveal that the knee of the curve of improvement is at about 10 foot candles and that beyond about 20 foot candles the gains are slight. It must be kept in mind that in measuring visual acuity, one is dealing with threshold values. It is highly questionable whether the almost microscopic gains in visual acuity obtained under the high foot candles justify their application to visual tasks where supra-threshold visibility is involved as in most everyday situations. Furthermore, data reveal that the visual acuity curve is practically horizontal from 50 foot candles to the higher levels.

Luckiesh and Moss (11) and Luckiesh (10) cite data on visual acuity for 1, 10, and 100 foot candles only. If they really desired to find the foot candle level beyond which no *practical* gains in visual acuity occur, they should have investigated the range between 10 and 100 foot can-

dles. As shown in Tinker's reviews (**29, 31**), this criticism may be aimed at all the basic data presented by Luckiesh (**10**). In some instances (decrease in heart rate, decrease in convergence reserve of ocular muscles), data for only 1 and 100 foot candles are presented. This procedure is inexcusable in experiments designed to determine how much light intensity is needed for efficient visual work. It appears, then, that visual acuity data are of only slight use for prescribing illumination intensities for visual discrimination in supra-threshold tasks. If accepted, there is no justification for suggesting that more than 40 to 50 foot candles are necessary for adequate discrimination even for tasks that approach threshold discrimination.

Visibility measurements. Luckiesh (**10**) states that "After establishing a standard of visibility or desirable see-level to be attained if possible for all tasks, it is seen that specifications of light and lighting and other aids to seeing can be based upon visibility measurements." The measurements are to be made by the Luckiesh-Moss Visibility Meter. This is a device consisting of two identical circular gradients which are rotated before the eyes to alter the brightness contrast of the object whose visibility is to be measured. It, therefore, reduces the object to threshold visibility. It is the threshold which is measured. Three assumptions are made: (a) Two objects are equal in visibility when both are barely visible, (b) "Two objects are equally above threshold visibility when their visibility has been increased by the same increase" in size, brightness, brightness contrast or time, (c) "The visibility of an object, or degree of supra-threshold visibility, is proportional to the decrease in any one of the fundamental factors necessary to reduce the object to threshold visibility." These assumptions are considered to be *axiomatic* and arguments against them are considered to be *futile*. Nevertheless, since recommended standards are based upon visibility measurements to a large degree, it seems desirable to examine the matter further. Things are not axiomatic just because some one says they are.

Since visibility measurements are in terms of threshold values, they are analogous to visual acuity measurements. They are subject, therefore, to the same criticisms as visual acuity measurements as criteria for prescribing illumination standards.

Luckiesh (**10**) emphasizes foot candles for equal visibility in prescribing illumination intensities. For example, to make newspaper text matter equivalent in visibility to 8 point book type on white paper under 10 foot candles of light, it is necessary to use 30 foot candles. And to make the 1/64″ divisions on a steel scale equal to this visibility level, 180 foot candles are needed. Are these levels of illumination intensity required for efficient and comfortable seeing? Luckiesh (**10**) assumes that this is a conservative standard. On his empirical scale, the 8 point type with 10 foot candles has 48 per cent maximum visibility. (Maxi-

mum visibility is obtained from a test-object whose critical detail has a visual size of 20 minutes; a critical detail of 1 minute is the smallest visible for persons with normal vision under 10 foot candles of light.) But no adequate experimental check is made for performance of these tasks under various levels of illumination. Tinker (27) found that the critical illumination level (the intensity beyond which no further change in reading performance occurs as the intensity is increased) for reading 7 point newspaper type to be approximately 7 foot candles. It is difficult to conceive the need of going above 20 foot candles to provide a margin of safety above the critical level. It is highly probable that an experimental check will reveal that other visual tasks, like discriminating the divisions on a steel scale, do not require the 180 foot candles indicated for efficient vision by the *computations* of Luckiesh. Related to this is the question of comfortable vision. Harrison (8), in discussing the difficulty of using high intensities because of the introduction of glare factors states "Visibility and comfort are two separate factors which do not always overlap completely."

No one will deny that visibility is an important factor in ease of seeing. But to prescribe standards in terms of scores derived from measurements made with the Visibility Meter is open to serious question. The basic data are threshold scores. While the derived scores may appear logical, supra-threshold seeing is not the same phenomenon as threshold seeing. Apparently, as illumination intensity is increased, one soon reaches a level of diminishing returns where further increase is of no practical importance or may introduce harmful factors from the viewpoint of easy and comfortable seeing.

Nervous muscular tension. Luckiesh and Moss (11, 12), place great stress upon the apparent decrease in nervous muscular tension during reading as the illumination intensity is increased from 1 to 10 to 100 foot candles. Tinker's (22) analysis of their data reveals that the method employed to present their results magnifies minute differences so that they appear large. Interpolation shows only gradual changes from 10 to 20 to 25 foot candles and very slight changes from there on to 100 foot candles. The conclusions that high foot candles are needed for ordinary reading is not valid. In a comparable situation, Tinker (25) found that for reading 10 point type, the critical intensity was about 3 foot candles. Below this level, rate of reading was retarded and fatigue increased, but for higher intensities there was no change. For people with normal vision, 10 to 15 foot candles should provide a satisfactory margin of safety for reading legible print.

Frequency of blinking. Another favorite criterion employed by Luckiesh and Moss (11, 12) and Luckiesh (10) as a basis for prescribing illumination intensities for visual work is frequency of blinking. The typical experiment is to measure the rate of involuntary blinking for the first and for the last five minutes for an hour's reading under 1, under

10 and under 100 foot candles of light. They note that the blink rate is greater under the 1 than under 10, and greater under 10 than under 100 foot candles. Therefore, it is concluded that relatively high intensities are desirable for reading. Even if these data are accepted as valid, we do not know where between 10 and 100 foot candles the curve of increased efficiency flattens out since intermediate intensity values were not studied. But there are several sources of information which suggest that blink rate is not a valid criterion of ease of seeing:

1. McFarland, Holway, and Hurvich (18), after a searching analysis of their own extensive experiments and of other studies, state: "A high blink-rate need mean neither an increase in fatigue nor an increase in difficulty of seeing." They conclude that "the rate of blinking can hardly be considered as a valid index of visual fatigue."

2. Tinker (32), in a study that has some bearing on the subject, found that frequency of blinking is an inadequate criterion of readability of print.

3. Bitterman (1), working with 3 and 91 foot candles of light, found that when subjects read for 40 minutes there was no significant difference in rate of blinking. In fact the frequency of blinking was slightly greater under the 91 foot candles. Incidentally, Bitterman also found no significant difference in blink rate for reading large type vs. small type. His results, therefore, indicate that rate of blinking cannot be employed as an index of ease of visual work. Further studies by Bitterman and Soloway (2, 3) showed that frequency of blinking is unrelated to duration of visual work or to the presence of a relatively intense glare source in the visual field. The reports of McNally (19) and Mac-Pherson (16) also cast doubt upon the validity of blinking as an index of ease of seeing.

4. The statistical treatment employed by Luckiesh and Moss (11, 12, 14) upon their data is open to severe criticism. Tinker (28, 29) has questioned the appropriateness of the geometric mean which they employ in most comparisons. The same criticism is raised by Hoffman (9). In a searching analysis, Hoffman also severely criticizes the use of the percentage technique employed by Luckiesh and Moss for presenting data, and for basing conclusions on percentage differences rather than on raw score differences. Percentage scores are notoriously unreliable. Furthermore, if the raw scores are below 100 (as most of them are), percentages magnify the differences. When percentages are used, therefore, the observed differences may be largely an effect of the derivation. Insignificant raw score differences may seem large when put into percentages. For instance, a typical average of 30 blinks during 5 minutes of reading is increased 10 per cent by a change of 3 blinks. Hoffman further points out that work decrement may be a more important variable than illumination changes in the results of Luckiesh and Moss. In general, he found little support for the contention that relatively high intensities are needed for effective and easy seeing.

5. Eames (5) criticizes Luckiesh and Moss (14) for using relatively few subjects in their experiments (including blink rate studies) and for employing "test wise" subjects. As pointed out by Eames, "People who take tests repeatedly in a given field gradually learn what is expected of them" and are un-

intentionally influenced by this knowledge. Results obtained under such conditions cannot be representative of the reactions of the general population.

The accumulated evidence indicates that rate of blinking cannot be accepted as a criterion for specifying intensities of light for visual work.

Decrease in heart rate. Luckiesh (10) and Luckiesh and Moss (11, 12, 14) cite data on change of heart rate while reading for one hour under 1 foot candle and under 100 foot candles of light. No data are presented for intermediate levels of illumination. It is stated that heart rate decreased 10 per cent under the 1 foot candle and 2 per cent under 100 foot candles. The conclusion was that from the viewpoint of ease of seeing the 100 foot candle level is desirable. An experiment by McFarland, Knehr and Berens (17) was designed to check the findings obtained in Luckiesh's laboratory. The results led to the conclusion that "It is questionable whether reliable criteria for determining adequate levels of illumination for tasks such as reading during short periods of time (approximately 2 hours) can be obtained in terms of . . . heart rate . . . ." Another check experiment was carried out by Bitterman (1), who recorded heart rate while reading under 3 and under 91 foot candles of light. "The results do not support the conclusions of Luckiesh and Moss with respect to the value of heart rate" as an index "of the ease of visual work." In view of the above evidence we must reject heart rate as a criterion for prescribing illumination intensities for visual work.

Decrease in convergence reserve. Luckiesh and Moss (11, 12, 14) and Luckiesh (10) cite data on decrease in convergence reserve of ocular muscles after reading for one hour under 1 and under 100 foot candles of light. The decrease was less under the 100 foot candles. No data are given for the range between 10 and 100 foot candles. We do not know, therefore, whether the 100 foot candles is significantly better than such levels as 20 or 30 foot candles.

Visual adaptation. Throughout their writings, Luckiesh and Moss (10, 11, 12, 14) emphasize that the eyes evolved under daylight levels of illumination and suggest the desirability of competing with daylight by artificial means. They consistently ignore the fact that the eyes readily adapt to easy and effective seeing over a wide range of illumination intensities.

Summary on intensity of illumination. Examination of the data employed by Luckiesh and Moss as a basis for specifying foot candle levels for visual work reveals a general lack of validity of these results as criteria for ease of seeing. The data from visual acuity, muscular tension and visibility measurements are misinterpreted or misapplied. The blink technique and rate of heart beat must be rejected because of lack of confirmation by independent workers. Furthermore the methods of

statistical analysis employed are frequently at fault. Any science of seeing based upon such an unstable foundation must, therefore, lack validity. Since these data have been the justification for specifying what appear to be excessively high levels of illumination intensity, we must reject such specifications unless justified by valid evidence from new experimentation.

LIGHTING CODES

School lighting. The American Recommended Practice of School Lighting (35) specifies the following minimum foot candles in service: 15 for classrooms, shops and offices; 25 for sewing and drafting rooms; and 30 for sight-saving classes. There is general agreement on the importance of hygienic illumination in reading and study situations. The recommended foot candle levels seem satisfactory in view of research findings other than those cited in the code. There should be, of course, a sound experimental basis for recommendations of this kind. Tinker (23) has pointed out that the recommended practice for school lighting is based upon conclusions derived from misinterpreted experimental results. Fortunately, the recommended practice is adequate in spite of inferences from inadequate data.

In a later publication by Sturrock (21), the foot candle levels are not in an approved code but are listed as the levels found desirable in the experience of successful business institutions, i.e., good present-day practice. For schools the foot candles listed include: 30 for study halls, class rooms, general laboratories, general manual training; 50 for drawing room, close work in laboratory, sight saving classes; 100 (considered especially low) for close work in manual training, and in sewing rooms. It is obvious to the impartial person who knows the field that these suggestions represent more intense illumination than is necessary for adequate seeing in the school situation. Data summarized by Tinker (24) and additional experimental evidence (25, 27) indicate that about 15 foot candles are adequate for ordinary schoolroom tasks and that 25 to 30 foot candles are satisfactory for the more severe tasks. Justification for the higher intensities is sought in the discussions of Luckiesh and Moss (12, 14) and Luckiesh (10). These have been evaluated above.

Office lighting. The Recommended Practice of Office Lighting (36) includes the following foot candle levels: 50 for difficult seeing tasks such as accounting, bookkeeping, and drafting; 25 for ordinary seeing tasks such as general office work, private office work, mail rooms; 10 for casual seeing tasks such as reception rooms and washrooms; 5 for simple seeing tasks such as halls and stairways. Considering the se-

verity of the tasks performed by some workers in general offices and special (as accounting) offices, the above recommendations are satisfactory. The 50 foot candles, however, should be considered liberal even for the difficult seeing tasks. The statement that "higher values will contribute greatly to accuracy, speed and ease" cannot be accepted as valid.

Sturrock's (21) summary of good present day practice does not deviate markedly from the recommended practice except that typing and prolonged reading of shorthand notes are listed at 50 foot candles and intermittent reading and writing at 30 foot candles. Each of these is about twice what is needed in terms of the visual task. The basis for the higher intensities is in terms of the discussions of Luckiesh and Moss (12, 14). The inadequacy of these data has been pointed out above.

Industrial lighting. A wide range of illumination intensities is recommended for various tasks in industry (37). Among the higher foot candle recommendations are: *over* 100 foot candles for such operations as extra fine assembly, automobile finishing and inspecting, cutting and sewing dark goods, engraving, proofreading, final inspection of tire casings, grading and sorting tobacco products, and certain inspection work in textiles; 50 to 100 foot candles for such operations as automobile assembly line, glass works inspection, fine inspection, bookkeeping, font assembly-sorting in printing industry, tin plate inspection, and stitching dark leather. With regard to all the recommendations, one is cautioned that the foot candles are minimum operating values and that in almost every instance higher values may be used with greater benefit.

It is stated that the recommendations are taken from a series of studies on the illumination needs of specific industries, or, if not available there, from current good practice. Examination of these studies (listed on page 23 of the report) indicates that in the main they are surveys rather than experiments. Furthermore, there is a lack of adequate descriptions of the survey techniques employed. In a few instances a general description of methods was given. Apparently what happened was first to make a survey of practice. This was followed by some sort of job analysis to determine what had to be discriminated. Then by reference to research studies (such as those reported by Luckiesh and Moss in their books) the intensity level of illumination presumably needed for the specific job was deduced. This method has some virtue provided sound data are referred to, which was not done in these cases. In a few instances it is stated that visibility measurements were made. Occasionally installations to achieve the recommendations

were made, the effect observed and additional modifications made. In no case was there experimental determination of the light intensity needed.

There are no valid experimental data which indicate that more than 50 foot candles are needed even for those practical visual tasks which approach threshold discrimination. Furthermore, as pointed out by Harrison (8), visual comfort may decrease under high intensities.

Home lighting. The most recent recommended practice for home lighting (38) specifies intensities ranging from 10 foot candles on card tables to 100 and more for sewing on dark goods. Forty foot candles are recommended for such situations as children's study table, kitchen work counter, laundry, and for prolonged reading. There is no valid reason for going above 25 to 30 foot candles for the more severe visual tasks in the home (24). Approximately 15 foot candles is adequate for many of these visual tasks. Figure 1 in *Recommended Practice of Home Lighting* (38) is misleading. "This chart shows the extent to which occupations and poor seeing conditions leave their mark on eyesight." The implication is that poor illumination causes ocular disability. There are no valid data which indicate this to be so. This chart represents an unjustified form of propaganda.

Present-day practice. Sturrock (21) has assembled foot candle levels of illumination which are labeled "good present-day practice." The tables are preceded by a classification (after Luckiesh and Moss) of foot candle needs for visual discrimination of tasks varying in difficulty. The material is apparently designed as a guide but is not necessarily in the form of recommendations. This sort of thing is valuable in many ways. But since it is based to a considerable degree upon the material presented by Luckiesh and Moss (12, 14) and by Luckiesh (10), the illumination intensities are excessively high in some instances—as 100 foot candles for sewing and proofreading, and 50 foot candles for reading small type and for kitchen counters. It should be pointed out, however, that much of the material is fairly satisfactory.

In general, recommended practice prior to 1940 (35) is fairly adequate, but as new codes are issued at later dates the apparent tendency has been to recommend as intense lighting as the traffic will bear. This is justified by referring to the work of engineers (largely Luckiesh and Moss) who state that these high intensities are nevertheless inadequate for easy seeing. As pointed out above, both the experiments and the conclusions which are cited as fundamental are frequently invalid. Furthermore the data are out of line with other independent experimental results.

VISUAL FACTORS

Eye disabilities. It is generally accepted that eyes with disabilities, even when corrected by glasses, need brighter light than normal eyes for adequate visual discrimination. Ferree and Rand (6) and Ferree, Rand and Lewis (7) are usually cited as supporting evidence. In the first study (6), it was found that apparent diopters of accommodation increased more for 14 presbyopes than for normal eyes in going from 1 to 5 to 25 foot candles of light. Interpolation indicates that for the normal eyes the curve of improvement shows little rise after about 8–10 foot candles; for the presbyopes, after about 15 foot candles. In addition, *one* myope and *one* presbyope were compared with a normal subject by measuring apparent diopters of accommodation at 13 intensities from 0.5 to 100 foot candles. The curve of efficiency for the normal person improved rapidly to 5 foot candles, then more slowly to about 20 and very gradually thereafter; for the myope there was considerable improvement to about 20 foot candles and little thereafter; for the presbyope there was considerable improvement to about 38 foot candles, and then slower improvement to 100 foot candles. It is of course impossible to generalize from *one case*, but apparently those with eye disabilities need somewhat brighter light than normals for clear seeing. This does not mean that they need 100 foot candles or more, as some people wish to imply.

In the other study (7) Ferree, Rand and Lewis were concerned with distant (20 feet) vision. The visual acuity for 4 presbyopes was compared with acuity for 3 normal people. The presbyopes continued to gain in visual acuity from 25 to 100 foot candles while the normal eye made little gain within this range. Since there is little or no relation between acuity of distant vision and acuity at near vision, these results have no bearing upon visual discrimination at the work surface (desk, work bench, etc.). Furthermore, one should not prescribe illumination for suprathreshold tasks in terms of threshold measurements (visual acuity). There is no evidence from these studies which implies that excessively high foot candles are necessary for those with ordinary visual disabilities. Rather, they suggest a moderate increase for those with corrected vision as compared with normal eyes.

Visual adaptation. It is well established that the eyes readily adapt to easy and effective seeing over a wide range of illumination intensities. This adaptation is rather slow in going from bright to dimmer illumination (for practical purposes, 15–20 minutes) and rapid in going from dim to bright illumination (1-3 minutes). Tinker (25) has demonstrated

that when adaptation is incomplete on shifting to a lower level of illumination, speed of perception is retarded. When adaptation is adequate, however, visual perception in reading is fully effective from 3 foot candles up for normal eyes in reading legible print. In another study, Tinker (26) showed that subjects tend to prefer for reading approximately the illumination intensity to which they have been adapted, whether it be 8 or 52 foot candles. These data indicate that readers tend to consider comfortable for easy reading any one of a wide range of illumination intensities provided such intensities are above critical levels and provided visual adaptation is adequate. Codes of lighting have consistently ignored the role of visual adaptation in seeing. They carefully point out that the eye has evolved under the bright illumination of daylight, but do not mention that the eye also evolved to see adequately at low as well as at high intensities of light.

ILLUMINATION FOR ADEQUATE SEEING

Critical levels of illumination. The critical level of illumination is the intensity beyond which there is no further increase in efficiency of performance as the foot candles become greater. Tinker (24) has summarized the data for critical levels of illumination: for reading of legible print (about 10 point on good paper) by adults, it is approximately 3 to 4 foot candles; for reading and study of children, 4 to 6 foot candles; for arithmetical computations, less than 9.6 foot candles; for sorting mail, 8 to 10 foot candles; for the exacting task of setting six-point type by hand, 20–22 foot candles; and for very fine discrimination required to thread a needle, 30 foot candles. In a later study, Tinker (27) found the critical level of illumination for reading newspaper print to be about 7 foot candles. Employing intensities from 2 to 55 foot candles, Rose and Rostas (20) found that reading efficiency, in terms of speed and comprehension, did not increase by a measurable amount with increased intensity of illumination.

, *Adequate levels of illumination.* It is obvious that visual work should not be done at critical levels of illumination. There should be an adequate margin of safety to provide for individual variation and the like. For such visual tasks as reading good-sized print (10 to 11 point) on a good quality paper, i.e., print of good legibility, 10 to 15 foot candles should provide hygienic conditions when one's eyes are normal. For situations comparable to the reading of newsprint, 15 to 20 foot candles should be adequate. In situations involving the reading of handwriting and other comparable tasks, 20 to 30 foot candles seem desirable. For

tasks comparable to discrimination of 6 point type, there should be 30 to 40 foot candles. And for the most severe tasks encountered in workday situations, 40 to 50 foot candles will be found adequate. There is no valid experimental evidence now available that indicates a need for over 50 foot candles intensity for adequate visual discrimination. The intensity values from 10 to 20 should be increased somewhat (5 to 10 foot candles) for eyes with slight disabilities or for those with corrections. For the higher values, however, no practical gain will be achieved for these people by increasing the intensity. The above suggestions hold for school children as well as for adults. In general, the child has much less severe visual tasks than adults.

Intensity of illumination cannot be prescribed without coordinating it with other factors such as distribution of light and brightness contrast. A good example of the uselessness of excessively bright light is found in the study by Darley and Ickis (4). They were concerned with vision in the drafting room, a very severe visual task. In comparing 30 with 75 foot candles of indirect light, they found the efficiency ratings for the two to be only slightly different. When they compared 40 with 80 foot candles of direct light (troffer) under conditions of no reflected glare, they also found no significant differences in the efficiency ratings. The observations of Harrison (8) are relevant here. He points out the danger of glare with installations of 50 foot candles and above of artificial illumination.

SUMMARY

Examination of the literature upon which lighting recommendations are based reveals that some techniques of experimentation are invalid, and that interpretations from certain other data are unwarranted. Some of the recommendations are adequate, others are not. The trend seems to be to specify as high intensities as the traffic will bear and at the same time to suggest to the consumer that if he uses still higher intensities he will improve his ease of seeing. All will agree that there should be sufficient light for adequate seeing. It is high time, however, that the consumer know what is adequate and what is surplus. As pointed out by Winslow (34), illumination should conform to real human needs. It is human health and comfort which are at stake.

In general the recommended practice concerning distribution of light, brightness contrast and color of light is satisfactory.

BIBLIOGRAPHY

1. BITTERMAN, M. E. Heart rate and frequency of blinking as indices of visual efficiency. *J. exp. Psychol.*, 1945, 35, 279–292.

2. BITTERMAN, M. E., & SOLOWAY, E. The relation between frequency of blinking and effort expended in mental work. *J. exp. Psychol.*, 1946, 36, 134–136.

3. BITTERMAN, M. E., & SOLOWAY, E. Frequency of blinking as a measure of visual efficiency: some methodological considerations. *Amer. J. Psychol.*, 1946, 59, 676–681.

4. DARLEY, W. G., & ICKIS, L. S. Lighting and seeing in the drafting room, *Illuminating Engineering*, 1941, 36, 1462–1487.

5. EAMES, T. H. Review of M. Luckiesh and F. K. Moss, *Reading as a visual task*, *Columbia Optometrist*, 1942, 16, 7.

6. FERREE, C. E., & RAND, GERTRUDE. The effect of intensity of illumination on the near point of vision and a comparison of the effect for presbyopic and non-presbyopic eyes. *Transactions I.E.S.*, 1933, 38, 590–611.

7. FERREE, C. E., RAND, GERTRUDE, & LEWIS, E. F. The effect of increase of intensity of light on the visual acuity of presbyopic and non-presbyopic eyes. *Transactions I.E.S.*, 1934, 29, 296–313.

8. HARRISON, W. What is wrong with our 50 foot-candle installations? *Transactions I.E.S.*, 1937, 32, 208–223.

9. HOFFMAN, A. C. Luckiesh and Moss on reading illumination. *J. appl. Psychol.*, 1947, 31, 44–53.

10. LUCKIESH, M. *Light, vision and seeing.* New York: Van Nostrand, 1944.

11. LUCKIESH, M., & Moss, F. K. *The new science of seeing.* Cleveland: General Electric Co., 1934.

12. LUCKIESH, M., & Moss, F. K. *The science of seeing.* New York: Van Nostrand, 1937.

13. LUCKIESH, M., & Moss, F. K. Illumination and eye health. *New Eng. J. Med.*, 1941, 224, 1117–1119.

14. LUCKIESH, M., & Moss, F. K. *Reading as a visual task.* New York: Van Nostrand, 1942.

15. LYTHGOE, R. J. *The measurement of visual acuity.* London: His Majesty's Stationery Office, 1932.

16. MACPHERSON, S. J. The effectiveness of lighting—its numerical assessment by methods based on blinking rate. *Transactions I.E.S.*, 1943, 38, 520–522.

17. McFARLAND, R. A., KNEHR, C. A., & BERENS, C. Metabolism and pulse rate as related to reading under high and low levels of illumination. *J. exp. Psychol.*, 1939, 25, 65–75.

18. McFARLAND, R. A., HOLWAY, A. H., & HURVICH, L. M. *Studies of visual fatigue.* Boston: Graduate School of Business Administration, Harvard Univ., 1942.

19. McNALLY, H. J. The readability of certain type sizes and forms in sight-saving classes. Teachers College, Columbia Univ., Contrib. to Educ., No. 883. New York: Bureau of Publications, Teachers College, Columbia Univ., 1943.

20. ROSE, F. C., & ROSTAS, S. M. The effect of illumination on reading rate and comprehension of college students. *J. educ. Psychol.*, 1946, 37, 279–292.

21. STURROCK, W. Levels of illumination. *Magazine of Light*, 1945, 14, No. 4, 26–36.

22. TINKER, M. A. Cautions concerning illumination intensities for reading. *Amer. J. Optom.*, 1935, 12, 43–51.

23. TINKER, M. A. The new "standards" for school lighting. *Sch. & Soc.*, 1939, 49, 95–96.

24. TINKER, M. A. Illumination standards for effective and comfortable vision. *J. consult. Psychol.*, 1939, **3**, 11–20.

25. TINKER, M. A. The effect of illumination intensities upon speed of perception and upon fatigue in reading. *J. educ. Psychol.*, 1939, **30**, 561–571.

26. TINKER, M. A. Effect of visual adaptation upon intensity of light preferred for reading. *Amer. J. Psychol.*, 1941, **54**, 559–563.

27. TINKER, M. A. Illumination intensities for reading newspaper typ. *J. educ. Psychol.*, 1943, **34**, 247–250.

28. TINKER, M. A. A reply to Dr. Luckiesh. *J. appl. Psychol.*, 1943, **27**, 469–472.

29. TINKER, M. A. Review of M. Luckiesh and F. K. Moss, *Reading as a visual task. J. appl. Psychol.*, 1943, **27**, 116–118.

30. TINKER, M. A. Illumination intensities preferred for reading with direct lighting. *Amer. J. Optom. & Arch. Amer. Acad. Optom.*, 1944, **21**, 213–219.

31. TINKER, M. A. Review of M. Luckiesh, *Light, vision and seeing. J. appl. Psychol.*, 1945, **29**, 252–253.

32. TINKER, M. A. Validity of frequency of blinking as a criterion of readability. *J. exp. Psychol.*, 1946, **36**, 453–460.

33. TINKER, M. A. Illumination standards. *Amer. J. Pub. Health*, 1946, **36**, 963–973.

34. WINSLOW, C. E. A. How many foot candles? *Amer. J. Pub. Health*, 1946, **36**, 1040–1041; also reprinted in *J. appl. Psychol.*, 1947, **31**, 140–142.

35. *American recommended practice of school lighting.* New York: Illuminating Engineering Soc. & Amer. Inst. Architects, 1938. Pp. 60.

36. *Recommended practice of office lighting.* New York: Illuminating Engineering Soc., 1942. Pp. 47.

37. *American recommended practice of industrial lighting.* New York: Illuminating Engineering Soc., 1942. Pp. 51.

38. *Recommended practice of home lighting.* New York: Illuminating Engineering Soc., 1945. Pp. 40.

ON FESTINGER'S EVALUATION OF SCALE ANALYSIS

LOUIS GUTTMAN

Department of Sociology and Anthropology, Cornell University

The theory of scale analysis had its origin some seven years ago. Since that time, especially by virtue of extensive and intensive research done in the Army, some of its further ramifications have been explored and several techniques have been devised for carrying an analysis out in practice. The power and incisiveness of this approach have been demonstrated in numerous attitude and opinion surveys made in the past several years, as well as in studies of achievement tests. A pleasing feature has been the simplicity of the techniques involved.

Most of the material, with respect to both applications and theoretical developments, is as yet unpublished. A manuscript has been prepared by Edward A. Suchman and the writer which will give the first comprehensive statement of both the theory and practice of scale analysis. This manuscript will form part of the four volumes soon to be published by the Social Science Research Council on the work of the Research Branch, Information and Education Division of the War Department. These volumes will also provide many illustrations of how scale analysis has been used for practical problems. Meanwhile, some brief statements of the principal concepts and instructions for practical procedures are available in article form to those who wish to use this approach in their own research (see the bibliography below).

On the basis of some articles which have been published and of some mimeographed progress reports, Festinger (1) has recently attempted a survey and evaluation of scale analysis. Since his survey is not based on all the information available, it is admittedly tentative and incomplete. In addition, full advantage has not been taken of the material which Festinger used as his sources; he raises a number of points which have already been answered there, and also introduces erroneous interpretations and conclusions.

It seems worthwhile to discuss at the present time some of Festinger's criticisms in order to help clarify the issues and to correct some important misapprehensions. Attention is also called to some articles that have appeared since Festinger prepared his paper, discussing various aspects of scale analysis (10, 11, 13).

Three of Festinger's points will be analyzed here: (a) criteria for scalability, (b) techniques of analysis, and (c) the use of scale analysis in practice. In the course of the discussion, some other aspects will be brought out which Festinger has not considered.

451

CRITERIA FOR SCALABILITY

Reproducibility. The main purpose of scale analysis is to test the hypothesis that a universe of qualitative items can be represented by a quantitative variable. In order for the universe to be represented exactly by a quantitative variable, each item must be a perfect function of that variable, or be perfectly reproducible from it. Thus the concept of reproducibility is paramount in scale analysis.

In practice, only a sample of items is used from the universe of content. Furthermore, in practice, it is not expected to find perfectly reproducible or scalable universes. Among other things, perfect reproducibility implies perfect test-retest reliability, which is certainly not to be expected empirically. However, if the reproducibility of the entire universe is very high, say over 90%, then that may be sufficient for many practical purposes. A quantitative variable which will represent an indefinitely large universe of items that well will ordinarily not lose much predictive power, whether used for predicting outside variables or whether predicted from outside variables. This will especially be true if the errors of reproducibility are random.

Since universe reproducibility must be estimated on the basis of only a sample of items, it becomes evident that the sample's reproducibility alone may not be a sufficient guide. Festinger criticizes the sample reproducibility coefficient for its inadequacy.* This inadequacy was recognized at the outset in scale analysis. The same kind of examples that Festinger uses (1, pp. 156–157), showing how five or nine statistically independent items can have high reproducibility, were worked out previously; several such examples will appear in the forthcoming volume. Indeed, there is an even worse case than that of statistical independence, namely that wherein some items have *negative* relationships with others; this is worse than being statistically independent from the point of view of scale analysis. Examples can be constructed showing how even in this case it is possible to have supriously high reproducibility in a small sample of items.

Festinger omits to point out that this problem about reproducibility was raised before, and that several answers have already come forth. In one of my mimeographed reports to which Festinger refers (15), there is the following question and answer:

Q. *Is reproducibility by itself a sufficient test of scalability?*
A. No. It is the principal test, but there are at least three other features

---

* Hausknecht (12) has raised this criticism earlier, also without taking cognizance of the fact that other criteria have always been used as discussed below.

that should be taken into account: (a) range of marginals, (b) random scatter of errors, (c) number of items in the sample.

Further questions and answers elaborate on the point. And again, in another paper (7) to which Festinger refers, it is stated:

The percent reproducibility alone is not sufficient to lead to the conclusion that the universe of content is scalable. The frequency of responses to each separate item must also be taken into account for a very simple reason. Reproducibility can be artificially high simply because one category in each item has a very high frequency. It can be proved that the reproducibility of an item can never be less than the largest frequency of its category, regardless of whether the area is scalable or not.

And further:

An empirical rule for judging the spuriousness of scale reproducibility has been adopted to be the following: No category should have more error in it than non-error.

If this latter rule alone were applied to Festinger's examples, it would immediately reject the hypothesis that the items are from scalable universes. The consideration about pattern of error would also disqualify the hypothesis that the items were from scalable universes.

An Alternative. One contribution to spuriously high reproducibility is the fact that each item is being related to a score which is based in part on the item. An alternative way to compute the coefficient of reproducibility is to hold out each item in turn from the analysis, thus obtaining N sets of trial scale scores. The errors for each item can then be counted from its relationship to the score based on the $N-1$ other items.

If this *partial-score* method were used on *statistically independent* items, then the reproducibility for each item would be precisely the relative frequency of its modal category. Thus, in Festinger's example (1, p. 156) of five independent dichotomies with marginals 80%, 60%, 50%, 40%, and 20% the respective modal relative frequencies are 80%, 60%, 50%, 60%, and 80%; hence, the reproducibility of all five items, computed from partial scores, would be the mean of the latter five percentages or 66%, compared with the spurious 86% Festinger obtained from whole scores. Indeed—*no matter what the interrelations of the five items were*—their reproducibility could not be less than 66%, because reproducibility of an item can never be less than its modal frequency. Similarly, in Festinger's second example (1, p. 157) of nine statistically independent dichotomies with marginals .9, .8, .7, .6, .5, .4, .3, .2, .1, the respective modal proportions of the items are .9, .8, .7, .6, .5, .6, .7, .8,

and .9, so the reproducibility of the set cannot be less than .72; Festinger finds .83 reproducibility from whole scores whereas if part scores were used the obtained reproducibility would be .72.

Items with extreme marginals like .9 and .1 do not help much in testing reproducibility since such items can never have more than 10% error.

In practice, it does not usually seem worthwhile to bother with partial scores, although this technique is available for doubtful cases. The fictitious examples of independent items do not illustrate what is to be expected in practice. Attitude (or achievement) items of the same general content are usually sufficiently correlated so that scores based on eleven of them will not be substantially different from scores based on twelve. Reproducibility from whole scores will not be much greater than from part scores—so their spurious excess of reproducibility over that from part scores can be largely ignored. Furthermore, even part-score reproducibility is not a sufficient test of scalability, for the additional criteria mentioned above must also be considered.

There is room for more improvement on criteria for scalability when samples of content are used, but it should be made clear that reproducibility by itself has not and is not the sole basis for drawing inferences from a sample of items. It is the basic one, because the reproducibility of the universe is essentially what is in question, but additional criteria have been and are being used.

Reliability. The suggestion that Festinger makes that the expected occurrence of scale responses be calculated under the assumption of a perfect scale plus a certain degree of unreliability is a promising one. This idea had been thought of in the earlier stages of the development of scale analysis but discarded in the form Festinger has suggested. The proportion of people with no scale errors *cannot* be properly calculated by the method that Festinger uses. Apparently he assumes that if .9 is the proportion of population responses that will be in the scale pattern for one question, then the proportion that will be jointly *within* the scale pattern for seven questions is $(.9)^7$ or 47.8%. Unfortunately, the same reasoning would say that the proportion of people who have seven responses *outside* the scale pattern should be $(.1)^7$; and in general the proportion of people with X scale responses and $7 - X$ scale errors should be given by the binomial distribution

$$\frac{7!}{X!(7 - X)!}(.9)^X(.1)^{7-X}.$$

But this is impossible, for nobody can have *all* his responses as scale

errors. Indeed, for the empirical example that Festinger borrowed (1, Fig. 2, p. 157), no matter what pattern of response a person may have, he can be placed into one of the scale patterns with at most four errors. Therefore, the range of possible errors for each person is 0 through 4, rather than 0 through 7 as Festinger supposes. This means that Festinger's calculations cannot be carried out consistently to estimate reproducibility under the given assumption. The difficulty is that whether a person will fall into the scale pattern is *not* independent of whether another of his responses is within the scale pattern. Unreliability does not behave that way with respect to the scale pattern.

The actual reproducibility of this example of seven questions was about .85 rather than the .9 Festinger assumed. It is interesting to note that $(.85)^7$ is .32, which is not far from the "over one-fourth" perfect scale types reported. Actually, the universe sampled by these seven questions would not now be accepted as sufficiently scalable but would be broken up into sub-universes; the study was made when 85% reproducibility was the empirical rule rather than the present 90%. The study did serve its purpose well, however, as collateral evidence presented there showed.

The further calculation that Festinger makes of adding 3.7% to his 47.8% seems based on an unfortunate double usage of the word "chance." In his second paragraph on p. 158 (1), "chance" is used to mean statistical independence between items. Such independence cannot exist simultaneously with the assumption of a scale pattern in his following paragraph; that is, the 7% who fall into perfect types under the hypothesis of independence of items have nothing to do with the distribution of error under the assumption of uni-dimensionality plus unreliability. The binomial distribution by itself—if it were correct— takes care of the second situation. Hence, Festinger's calculations are incompatible in adding 3.7% (7% of $1 - .478$) to 47.8% to obtain 52.2% as the "chance" proportion. The 7% is correct for independent items; the 47.8% would be correct for the scale-plus unreliability case if the binomial hypothesis held; and the two cases do not hold simultaneously. "Chance" means something different in each case.

A consistent use of reliability. Several correct approaches to the use of the concept of unreliability are possible, instead of the inconsistent binomial approach. One such approach will be sketched here briefly for the case of dichotomies. Let n be the number of dichotomies in the sample of items so that there are $n+1$ scale types or ranks possible. Let r be the rank of the type that is "positive" on r of the items; r ranges from 0 to n. Let Pr be the proportion of the population whose

"true" rank on the n items is r, and let Prj be the probability a person of "true" rank r will be "positive" in the jth item ($j = 1, 2, \ldots, n$). There are 2^n types of people—scale and non-scale—possible on the n dichotomies. The expected proportion in each of the 2^n types can be calculated from the Pr and Prj. Conversely, from the observed 2^n proportions in an actual experiment, the Pr and Prj can be estimated. There are $n+1$ parameters Pr, of which n are independent since their sum must be unity. There are $(n+1)n$ parameters Prj, all of which are independent. Hence, there are $n+(n+1)n$, or $n(n+2)$, independent parameters to be estimated from 2^n-1 independent observations. If n is greater than 5, this provides more equations than there are unknowns—*so the hypothesis of the scale structure can be tested*, as well as having the parameters estimated. Unfortunately, the equations involved in the above analysis are curvilinear, and do not seem to lend themselves to practical use because of the difficulties in the numerical computations. Furthermore, even this analysis has been simplified by assuming that persons within the same "true" rank were equally reliable within each item. Without this simplifying assumption, the equations would have innumerably more parameters.

In any analysis using the concept of test-retest reliability, it must be remembered that scalable data must in general be highly reliable, although the converse is not necessarily true. The coefficient of reproducibility—especially if computed by the part-score technique described above—sets a *lower bound* to the average reliabilities of the items (6, and especially 8). In particular, if items are perfectly reproducible, they are perfectly reliable. Hence, Festinger errs in his assertion that "Even if a perfect scale were achieved these claims [concerning invariance properties] would all be limited by the degree of reliability . . . of the questions asked" (1, p. 160). Perfectly scalable data are perforce perfectly reliable. Conversely, highly unreliable data cannot be scalable. One of the contributions of a scale analysis is to provide automatically information about reliability by helping set a lower bound to it for each item.

The simple criteria used in conjunction with that of reproducibility for sample data do serve to distinguish between data that are highly scalable and those that are not. The case where the items are independent will always be rejected on the basis merely of the criterion of improvement, namely, that no category should have more errors than non-errors. The further criteria of studying patterns of error also tend to insure that no dominant second variable is present even if reproducibility is high. That is what is meant by the statement that "in imperfect

scales, scale analysis picks out deviants or non-scale types for case studies." If no non-scale types have substantial frequencies, then that tends to indicate that there is no substantial second factor present. However, if one or more non-scale types do have a substantial frequency, then that is an indication of where an additional factor (or factors) is entering into the picture. If an additional factor is sufficiently prominent, it may be worthwhile to try to piece it out further by asking additional questions. The universe might be divided into two or more sub-universes, each of which may be scalable separately. Or it may turn out that the additional factor is so highly correlated with the most dominant factor that it does not make much difference whether they are treated as two separate variables or as a single variable.

The problem is not to find out whether a perfect scale is present in practice, but rather whether it is worth worrying about any additional variables that may be present. The criteria used in practice are believed to provide an answer to this and to decide properly whether or not a set of data can be regarded as sufficiently scalable for most practical purposes.

Quasi-scales. One kind of non-scalable universe is called a *quasi-scale*. A quasi-scale is different from a scale, not just in the reproducibility, but in *the entire pattern of responses*. Festinger seems to have misunderstood the definition of a quasi-scale, for he seems to believe that it differs from a scale only with respect to reproducibility (1, p. 156 and p. 159). A universe which is quasi-scalable will ordinarily have less than 85% or so reproducibility, but that is not its distinguishing feature. The distinguishing feature is *the gradient in the responses to the items*. Cutting points cannot be established (as in the case of a scale) which will enable one to say that a person above the point is in one category of an item and a person below the point is in another category; but one can state that, if one person is higher than another in the quasi-scale, then his probability of being in a higher category of an item is correspondingly greater.

There are many kinds of configuration which are less than 85% or 90% reproducible and which are not quasi-scales at all. For example, an area may have two or more dominant factors in it, in which case it would not be either a scale or a quasi-scale. In a quasi-scale, there are one dominant factor and infinitely many small factors. The order of people in a quasi-scale is according to the dominant factor, and is essentially invariant from sample of items to sample of items, provided that the samples are large enough. There is a great deal of work yet to be done on the theory of quasi-scales, but enough is known to say that

they have quite a different character from scales and from other kinds of universes. Another distinguishing feature between a scale and a quasi-scale is that the scale has an intensity function and further meaningful components, whereas a quasi-scale does not have an intensity function or further components of that kind.

Neurotic phenomena have been found to be quasi-scalable. For example, the Neuro-Psychiatric Screening Adjunct, which is the official paper and pencil test used at all military stations since October, 1944, is a quasi-scale and is a product of a rigorous investigation of efficient screening tests made possible by the scale analysis approach (16, 17).

TECHNIQUES FOR SCALE ANALYSIS

Scalogram devices. There are several alternative schemes now available by which to carry out a scale analysis in practice. They are virtually equivalent in terms of the results they yield, but they differ somewhat in operation. *Scalogram boards* have been the principal device used by the War Department, and are perhaps the most flexible and easiest to use. The boards are relatively simple to make and to operate; the cost depends upon how large a board is desired and whether or not a pair is to be made. If a single board is used instead of two, then the workmanship need not be precise and the board can be made fairly cheaply by any carpenter. There are alternative mechanical schemes that might be used instead of the wooden board, and undoubtedly other schemes will be invented in the future which will be even easier to construct. Instructions for the construction and use of a scalogram board will appear in the forthcoming volumes on the work of the Research Branch.

The Cornell technique (7) is also very easy to learn; it is taught in a course on attitude and public opinion analysis to students who have no background whatsoever in statistics. For achievement tests, where all items are dichotomous—being marked either right or wrong—the Cornell technique is perhaps the best of all to be used. For dichotomies, there is no problem of combination of categories, so that there is but one trial to be made in an analysis. The Cornell technique suffers a bit in flexibility compared to the scalogram board when a series of trials has to be made. Ordinarily, but two trials may be needed in an analysis, and the Cornell technique has proved very advantageous in such cases for general research purposes. It can be carried out on IBM equipment as well as by hand.

The Goodenough technique (2) is based upon an explicit tabulation of all combinations of responses that actually occur. It is more "rigor-

ous" than the preceding two techniques in that it counts the errors at each stage. However, it yields no different results in the end. Apparently Festinger has not worked through the Goodenough technique to see how it does work out in practice.* The first step seems simple, but it takes a good deal of experience to master the three following steps. The process becomes very bulky and involved when ten or twelve items are used.

The Cornell Technique has the advantage that its complexity does not at all change, regardless of the number of items (though of course the amount of labor increases with the number of items). The same lack of increase of complexity holds to a slightly less degree with the scalogram board.

The problem of metric. The earliest technique for scale analysis was that of least squares (3). It is quite properly to be abandoned as a procedure in practice because it is certainly far more cumbersome than the others. However, the equations involved have turned out to be of basic importance in *interpreting* a scale, and have led in particular to the empirical treatment of the intensity function which is proving so vital for attitude and public opinion work. Also, the basic thinking behind the equations have led to a solution to the related problem of paired-comparisons (12).

In the beginning of my work on scale analysis, I had thought that one of the most important problems was that of metric. I had thought that how to obtain weights for items was perhaps the leading problem to be solved. But as the theory of scale analysis developed, it became clear that the problem of weights was essentially a minor one for most practical purposes. Indeed, for the *perfect* scale pattern, it is easy to see that if scores are to be obtained for people by adding up weights assigned to categories of items, then, no matter what weights are used—as long as they have the proper rank order within each item—the scores of the people will have *exactly* the same rank order. The ordering of people in this sense does not depend at all upon finding a particular weighting system.

The important problem turned out to be that of finding the *structure* a universe of items must have in order to be scalable; it was not that of finding weights.

* Festinger also apparently has misread Goodenough as to how to measure reproductibility. Goodenough explicitly says that "at least 85% of the total number of responses must fall within the scale pattern, so that it is possible to reproduce 85% correctly all the responses of all the respondents from the scale scores" (2, p. 184). Festinger seems to have misread this to mean that 85% of the individuals fall into perfect scale types.

The problem of a metric does turn out to come into the picture for further problems, and it first appeared as a practical problem with respect to that of bias in questionnaire wording (11, 13, 14). The problem here was, after people are ranked from a high to a low on an attitude or opinion, to find a dividing point in the order such that the people on one side can be called positive and people on the other side can be called negative. The equations of scale analysis, when applied to the perfect scale pattern, show a most remarkable result. They show that a universe of items which is perfectly scalable can be resolved into an infinite series of principal components, the first of which provides the basic metric, the second of which is the intensity component, and the remaining ones are as yet not named (10). Empirical study of the intensity function has afforded for the first time a scientific solution to the problem of question bias.

These equations, then, show that a scalable attitude is somewhat different from the twelve-inch ruler that Festinger uses as an analogy (1, p. 160). The responses of a person to items in a scalable universe are seen by means of these equations to be a function of the person's metric score, his intensity, and the further components in the scale. The person's rank order is sufficient to reproduce his responses exactly; in this sense, the responses of the population are but a function of a single variable. Resolving the responses into components by the alternative device of the least squares equations shows the responses to be a function of infinitely many variables, each of which is a function of the rank order.

These striking results from using the least squares equations in conjunction with the perfect scale pattern will be elaborated on in the forthcoming publication on the work of the Research Branch. It might further be pointed out here that these equations resolve also the paradox which appears in achievement tests where the *difficulty* of an item seems to introduce a factor different from the common *content* factor that the items may have. Since scale analysis applies to achievement tests as well as to attitude or opinion areas, achievement tests also are resolvable into the principal components of a scale. In a scalable achievement test, then, each item is a function of but a single dimension from the point of view of reproducibility, but a function of infinitely many dimensions from the point of view of principal components. The apparent contradiction between these two points of view is resolved by the fact that the infinitely many principal components in turn are perfect functions of the rank order of people.

Uses of Scale Analysis

Incidence of scales. The theory and techniques of scale analysis provide a test of the hypothesis that a universe of qualitative items can be represented by a single quantitative variable. This hypothesis is appropriate for any qualitative universe obtained by any method of observation. The universe may be a set of items recorded on a questionnaire, or observations obtained in non-directive interviews, by participant observation, or by any other technique of gathering data. No matter how the data are gathered, each observation is but a sample of all similar observations that could have been obtained, and the entire universe of observations is ordinarily of interest.

As Festinger suggests, scalable universes may be the exception rather than the rule. Festinger does not give any explicit reasons for his belief, but this position will be substantiated in the forthcoming volume. It has already been pointed out that one possible reason for the existence of an attitude scale is that of a homogeneous culture (4, p. 149). If a population is not subjected to the same social stimuli with respect to the attitude, it might be expected that it will prove to be unscalable for them. The fact that neurotic phenomena have not been found scalable can perhaps be explained in this fashion. Similarly, an area of achievement may be expected not to be scalable if there is no uniform program of training for the population involved.

Another reason for expecting many universes not to be scalable in practice is that the notion of a universe is so comprehensive. Each sub-universe of a universe is of course itself a universe. Since there is ordinarily a vast number of imperfectly related sub-universes, there must be a vast number of combinations of them which are non-scalable universes. Merely this formal consideration would lead one to believe that most universes are not scalable. Non-scalable universes may of course be broken down in some cases into scalable sub-universes. One of the contributions of scale analysis is to point out the need for being clear about the universe's content. By focusing on more and more homogeneous content, research can be made more meaningful and external predictions be made more effective in the long run.

The development of the above-mentioned screening test for psychoneurotics (16, 17) is but one example of how research utilizing scale analysis was more effective than it would have been had the more traditional but less incisive procedures been followed. Instead of throwing together all kinds of conceivable predictive items into one composite, fifteen different universes of content were defined which might be re-

lated to the criterion of psychoneuroticism. The structure of each of these universes was first analyzed separately. Because each was found to be either a scale or a quasi-scale, only a relatively few items from each were needed in order fully to utilize the predictive power of the universes. The multiple correlation of the criterion was then worked out on all fifteen predictors with the finding that one of the universes predicted as well as the best combination of the fifteen. This enabled the short but efficient screening test to be used with the knowledge that it retained the predictive power of innumerably many items in fifteen different universes. Such a complete usage of predictive power could not have been made without scale analysis.

From the practical point of view, another important feature here is the amount of labor saved by scalogram techniques in obtaining this maximum predictive power, compared to using more traditional techniques which are far more laborious and which would yield less effective predictions.

The two problems, that of scalability and that of external prediction are distinct but related. By focusing on the scaling problem in its own right, more effective external predictions are thereby made possible.

There are many areas which have been found to be scalable thus far, and therefore these areas can be handled economically by means of simple scale scores. Many areas have also been found not to be scalable; all such areas cannot be handled so simply. It is known how to treat quasi-scalable areas, and Lazarsfeld is now completing a theory of the latent dichotomy which also can be handled by means of a single quantification. How to utilize other kinds of non-scalable areas is still an unsolved problem. The emphasis that scale analysis makes in this connection is that unless the structure of the universe is known, it is not known how best to treat the universe for any particular purpose.

Distinction between theory and techniques. The basic *theory* of scale analysis is not to be confused with particular *techniques* for carrying out such an analysis in various kinds of situations. Festinger borders on confusing the two when he states that " 'scale analysis' seems to be an excellent technique for use with paper and pencil tests or other instances of measurement where the situation permits the inclusion of several questions centering about the same topic" (1, p. 160). If a research problem is concerned with a universe of content, then that universe must be studied. That is what the theory calls for. How adequate is the technique which Festinger implicitly advocates of studying only a single item from the universe?

One of the important aspects of a universe of content is its structure;

for example, is the universe scalable or does it have some other kind of structure? The theory of scale analysis tells what a scalable structure is, and the various properties possessed by such a structure.

The practical problem is to obtain information about the structure from only a sample of items. It has already been indicated how an adequate sample of items can be chosen to test the hypothesis of scalability. Furthermore, the number of items to be used in a pretest must be distinguished from the number of items to be used in a final study. One of the properties of a scalable universe is that only one or two items can be used in a final study for many purposes once their place in the universe is ascertained. The scalability of the universe must first be analyzed, however, by a dozen or so items in a pretest.

The statement that "most of those engaged in this type of research [public opinion] will probably find the inclusion of a series of questions which could be subjected to scale analysis not feasible from practical considerations" (1, p. 159) does not accord with what is the actual practice both in public opinion and in market research, as well as in general attitude research. It is because workers in these fields are concerned with a universe of content that they pretest various questions on the same topic; it is a foolhardy pollster who bases conclusions on but a single question. The use of the split-ballot is evidence of this concern with sampling of content. In addition, ordinary polls often include several questions on the same topic on the same ballot. The extreme position taken by advocates of "open-ended interviewing" is to ask a whole series of questions of every respondent. And of course, conventional attitude surveys almost invariably use a substantial set of questions for a given topic.

It is a misapprehension to believe that asking several questions on the same topic necessarily creates a problem of rapport. In one survey made of a national cross-section by a leading public opinion polling agency, an area of content was defined and then sampled by four questions. Some of the interviewers complained because of the great similarity of wording of questions. The questions were very similarly worded because the content concerned the size of the Navy and was very hard to discuss in different ways. But even under these adverse circumstances, the analysis was successful in showing that the area was scalable and that the zero point could be located properly by the intensity function. Even more questions in the same area had been used in the pretest in Ithaca on a cross-section of the population there, and interestingly enough there was no complaint either from the respondents or from the interviewers, although the interviewers were no different from those

used in the national cross-section and had no knowledge whatsoever of what was involved in scale analysis. An area of apparently very similar questions is an exception rather than the rule. The example about desire for post-war schooling that Festinger has borrowed (1, p. 157) certainly provides no problem of rapport, and the general run of areas studied by public opinion polls do not present any particular problem of rapport. Another large market research agency has tried scale analysis in a routine study and has found no difficulty whatsoever with it. Because of its simplicity and its objective solution to the problem of bias, this agency plans to use this approach regularly.

It seems premature, then, to conclude that scale analysis cannot be carried out in practice in public opinion work. To the contrary, scale analysis is becoming more essential in this field because it affords for the first time a scientific solution to the basic problem of bias in public opinion polls. This problem arises from the fact that a universe of content is being studied and any single question is but a sample of all possible questions that could have been asked. How can one determine which question does coincide with the zero point of the entire universe, that is, the point which divides those who are negative on the issue from those who are positive?

The intensity function provides a scientific solution to this problem (13). It provides both a definition and a technique for ascertaining a zero point for the population. Unless some such objective approach to the question of bias is used in public opinion polls, it cannot be certain how much credence to place on their reports.

By providing a solution to the problem of bias, scale analysis clears the way for asking questions in the manner which will best help establish rapport with the respondent. The particular form of a question does not affect the results of scale analysis, so the research worker can concentrate on obtaining the wording which will make the interviewing work go most smoothly. Thus scale analysis has a contribution to make toward increasing rapport in surveys rather than the contrary. Apprehension that the opposite is true seems to be due to a misconception that scale analysis presupposes a particular way of asking questions.

If progress is to be made in the scientific study of attitudes, public opinion, and achievement, it seems necessary to concentrate on the problem of the structure of content. Techniques are not worth much if not guided by any thoery. The theory of scale analysis happens to lend to simple and practical techniques. To compare these techniques with others, one would have to ask: what theory of structure guides the alternative techniques and how adequately is this theory served thereby?

BIBLIOGRAPHY

1. FESTINGER, L. The treatment of qualitative data by "scale analysis." *Psychol. Bull.*, 1947, **44**, 149–161.

2. GOODENOUGH, W. H. A technique for scale analysis. *Educ. psychol. Msmt.*, 1944, **4**, 179–190.

3. GUTTMAN, L. The quantification of a class of attributes: A theory and method of scale construction. In P. Horst *et al.*, *The prediction of personal adjustment*, Soc. Sci. Res. Council Bull. No. 48, 1941, 319–348.

4. GUTTMAN, L. A basis for scaling qualitative data. *Amer. sociol. Rev.*, 1944, **9**, 139–150.

5. GUTTMAN, L. Scale and intensity analysis for attitude, opinion, and achievement. (Mimeographed, 1945. To appear in the proceedings of the conference on military contributions to methodology in applied psychology to be published by the Univ. of Maryland Press.)

6. GUTTMAN, L. A basis for analyzing test-retest reliability. *Psychometrika*, 1945, **10**, 255–282.

7. GUTTMAN, L. The Cornell technique for scale and intensity analysis. (Mimeographed, 1946. To appear in the proceedings of the conference on measurement of consumer interest to be published by the Univ. of Pennsylvania Press; also in *Educ. psychol. Msmt.*)

8. GUTTMAN, L. The test-retest reliability of qualitative data. *Psychometrika*, 1946, **11**, 81–95.

9. GUTTMAN, L. An approach for quantifying paired comparisons and rank order. *Ann. math. Statist.*, 1946, **17**, 144–163.

10. GUTTMAN, L. Suggestions for further research in scale and intensity analysis of attitudes and opinions. *Int. J. Opin. & Att. Res.*, 1947, **1**, 30–35.

11. GUTTMAN, L., & SUCHMAN, E. A. Intensity and a zero point for attitude analysis. *Amer. sociol. Rev.*, 1947, **12**, 57–67.

12. HAUSKNECHT, G. A procedure for determining a useful approximation to an ideal scale. Unpublished manuscript.

13. SUCHMAN, E. A., & GUTTMAN, L. A solution to the problem of question bias. *Publ. Opin. Quart.*, 1947. (In Press.)

14. *Experiments on the measurement of the intensity function and zero point in attitude analysis.* Research Branch, Information and Education Division, Army Service Forces. Report D-1, Mimeographed, 1945.

15. *Questions and answers about scale analysis.* Research Branch, Information and Education Division, Army Service Forces. Report D-2, Mimeographed, 1945.

16. *The screening of neurotics.* Research Branch, Information and Education Division, Army Service Forces, Report B-104, 1944.

17. *A study of psychoneurotics in the army.* Research Branch, Information and Education Division, Army Service Forces. Report B-107, 1944.

NOTE ON "A REVIEW OF LEADERSHIP STUDIES WITH PARTICULAR REFERENCE TO MILITARY PROBLEMS"[1]

DONALD E. BAIER

Personnel Research Section, A.G.O.

The valuable report[2] with which this note is concerned "... summarizes and reviews selected references from the available literature dealing with the problem of the selection of leaders in various fields. The primary interest in preparing the article was to provide a summary of techniques and results that would be of value to psychologists dealing with problems of selecting leaders, particularly in the military field."

It is the purpose of this note to make available additional facts and comments which appear to bear on the following conclusions of the reviewer:

1. "Progress has not been made in the development of criteria of leadership behavior . . . ."

2. "Advances in methodology in this field are definitely not striking."

It is this writer's belief that these conclusions, insofar as they are meant to apply to military leadership, are not entirely warranted.

In two reports[3] published by the Medical Field Research Laboratory, Camp Le Jeune, N. C., research on measurement of "leadership" is reported. These studies indicate a substantial relationship (tetrachoric $r = .42$) between superior officers' reports of the combat performance of Marine Corps officers graduated from the Corps Officer Candidate School and the standing of these graduates among their fellow-marines as indicated by a nomination procedure conducted during their pre-officer training. The two sets of evaluations were completely independent.

An as yet unpublished follow-up study by the Personnel Research Section, AGO, of West Point graduates after 18 months of duty as Army officers also reveals a significant association ($r = .51$ for Infantry Officers) between inter-cadet ratings or leader-nominations and success as an officer measured by the *Officer Efficiency Report, WD, AGO Form 67*. Here again there is basis for believing that the two measures are independent.

---

[1] The opinions expressed herein are those of the author and do not necessarily represent the official view of the War Department.

[2] JENKINS, WILLIAM O. A review of leadership studies with particular reference to military problems. *Psychol. Bull.*, 1947, 44, 54–79.

[3] *Validation of officer selection tests by means of combat proficiency ratings.* Medical Field Research Laboratory Report No. 1, January 18, 1946 and No. 2, May 16, 1946. Camp Le Jeune, N. C.

The reviewer's account of the research upon which are based the current methods for selecting wartime officers for integration into the regular Army may result in misunderstanding. In discussing the correlation between the Officer Evaluation Report and the criterion of leadership, the latter being a product of nominations by subordinates and peers with a veto power resting with the commanding officer of the group, Jenkins states:

> . . . The degree to which the Commanding Officers' ratings were weighted in the Officer Evaluation Report was not stated, but it appears likely that this factor played an important role. Substantial agreement between ratings by the C.O. and by fellow officers was to be expected. Since the OER had the highest validity, and the other measures when combined with it increased its correlation with the criterion only .07, these questions suggest the necessity for a further examination of the nature of the criterion here employed (p. 74).

The Officer Evaluation Report was accomplished in the majority of cases by the immediate supervisor, not the C.O., and represented only the former's evaluation of the ratee. The conclusion that substantial agreement between ratings by the C.O. and by fellow officers was to be expected does not appear to be justified. The C.O. was only one of from 7 to 30 nominators who participated in determining the ratee's criterion standing. He had no knowledge of how the other members of the nomination group evaluated each ratee, and his rating was used only to eliminate from the criterion groups of High, Low or Middle those rare cases where the C.O. placed the rated officer in the opposite extreme from the combined ratings of his subordinates and peers. Later studies employing a criterion which did not include the C.O. showed no drop in the validity of the OER or FCL type rating device. It is our belief that the nomination criterion employed in the studies cited does represent progress in the development of leadership criteria.

With respect to methodology, the forced choice technique as exemplified in the triads and tetrads of the OER and the recently revised Army Officer Efficiency Report seems to deserve more attention than the reviewer accords it. This technique, which has been described briefly in a paper titled "The Forced Choice Technique and Rating Scales," presented at the American Psychological Association meeting in Philadelphia on 5 Sept. 1946 by the Personnel Research Section, AGO, not only provides valid indicators of the ratees' standing on a nomination criterion, but favorably influences ratings of overall competence (if they are made immediately following completion of the FCL items) so that they show substantially less negative skewness. Clearly, the forced-choice technique is effective in diminishing rater-bias and in improving the distribution and validity of ratings which are generally regarded as indicative of leadership performance.

BOOK REVIEWS

MUNN, NORMAN L. *Psychology: The fundamentals of human adjustment.*
Boston: Houghton Mifflin, 1946. Pp. xviii+497.

The importance of the introductory course in psychology cannot be
overestimated for it determines to a great extent the student's attitude
toward the subject and whether or not he goes any further with it. But
the importance of the textbook depends to a large degree upon the in-
structor. Some instructors lean heavily upon the text, others hardly
at all. In reviewing a book, however, evaluation must be made as if it
were the sole source of the student's introduction to the subject, re-
gardless of the instructor's predilections, interpretations, choice of ma-
terial, or method of handling the course. Though there are suggested
readings at the end of each chapter in this as in other texts which the
student is urged to consult, their influence is admittedly minor since the
author of a text, as Munn says, writes with the feeling that he, in com-
mon with most teachers of the subject, could "organize its topics in a
more logical sequence, choose apter illustrations, find more interesting
examples and . . . write a book that . . . would be more appealing to
instructors and students than any he has seen" (ii). Some other re-
quirements of a good introductory text are succinctly suggested by
President Leonard Carmichael in his Introduction wherein he discusses
the reasons for studying Psychology today: as an essential part of a
general education; as preparation for professions like law, medicine,
teaching, the ministry, and business; and for further professional work
in the subject itself. That Munn has met both his own and the editor's
demands with considerable success there can be no question. The book
is plentifully provided with excellent diagrams, half-tones, and tabu-
lar matter; it is full of concrete material chosen from a wide variety of
sources; and its approach is scientific throughout.

It would seem, in the light of these virtues, that this text meets the
requirements of the introductory course almost to perfection. It is such
a splendid job in so many respects that any criticism at all seems super-
erogatory, if not hypercritical; and yet there are qualities expected in an
elementary text which are of equal or greater importance than the ones
which this book has in such large measure. The most important lack is
an underlying point of view or theoretical structure integrating and
unifying the topics and their relations to each other and the subject as a
whole. We shall find that in this respect the book is not up to its ac-
complishments otherwise.

The book is divided into seven main parts and these in turn into two
or more chapters. It begins with a discussion of general, methodological
and historical material, then proceeds to consider in turn the anatomical
and physiological bases of behavior, learning, remembering, thinking,

468

motivation, conflict, feeling, perceiving, the special senses, statistics, intelligence, aptitudes, and personality. The treatment develops by consideration of simpler processes followed by the more complex, in so far as possible, though there is some back-tracking which is done with a minimum of repetition of earlier material. With the general plan of the work before us we can now consider it more in detail.

Beginning with the origin and scope of psychology, the first two chapters are devoted to a brief glance at the history of the subject through a consideration of such topics as the psyche, the organism, methods in philosophy, physiology, and physics, analysis of consciousness and some fields of psychology. Chapters 1 and 2 really constitute a single topic or set of topics and furnish an excellent survey of methods, fields, and problems. They are properly brief, to the point, and very readable. Only in one detail does the text here need emendation. In discussing scientific controls it is stated that "there is never more than one independent variable in a given experiment . . . . If two or more factors were varied, he (the scientist) obviously would not know which had produced the phenomena observed" (p. 23). While it is not expected that the logic of analysis of variance and designed experiments should be presented at the elementary level (though it is not impossible) advances in statistics have made the older Mill-Bacon canons of scientific procedure represented by this statement quite out-of-date. Variation of only a single variable in psychological experiments is possible so seldom as to be almost a fiction and now that we have the statistical tools for handling multiple variates we might as well give up the fiction.

Part 2 deals with psychological development and consists of three chapters: origin and psychological significance of response systems, conception to maturity, and factors in psychological growth. Here the biological bases of behavior are explained and the psychological processes most directly correlated with them are brought in. The result is that the simpler and more complex processes are more or less intermingled in these chapters as a partial list of the topics reveals: structure and functions of receptor and nervous systems, embryonic development, sensitivity, locomotion, prehension, language, gestures, writing, genes, heredity, environment, and maturation. The order is in general from simple to complex but there are some reversals. Thus one would expect a discussion of genes and embryological development before discussion of the nervous system but here it follows the latter. The reason for Munn's order is obvious to a reader of the book and a good one: discussion of the more elementary gene units links directly with problems of heredity, environment, maturation and growth. There is no discussion of nerve action potentials and the treatment of the autonomic nervous system is postponed to the chapter on emotion where diagrams illustrating its relations to the cerebro-spinal system are given. The reviewer finds it impossible to omit the autonomic nervous system when

explaining the rest of the nervous system, although like Munn, he finds greatest use for it in the discussion of feelings and emotions. Figure 13 in chapter 3, showing the spinal reflex-arc system, would not be over-complicated if the sympathetic ganglion and its fibers were included as is usually done, and with some textual discussion this would remedy a serious omission at this point. On the other hand, Munn has included more material on the nervous system than is generally presented. The diagrams showing different types of synaptic connections make inter-action, facilitation, and inhibition intelligible neurologically. The dis-cussion of cortical representation of sensory functions is especially well done.

The chapters on conditioning, learning, memory and thinking suc-ceed in presenting a considerable amount of material but suffer from the lack of clear integrating principles. While Munn rejects classical, Pavlovian conditioning theory as an adequate account of all learned behavior it is not clear what principles he would employ instead. That the author places greatest reliance on trial and error, past experience, and association appears from his treatment of certain particular prob-lems rather than from explicit structuration of the material. One must dig his underlying approach out from a few critical cases which reveal the author's fundamental position. Thus the explanation of how the chimpanzee reaches an apparently inaccessible object is a case in point. According to Munn we are to suppose " . . . a chimpanzee has, in the jungle, learned to reach otherwise inaccessible objects by swinging toward them on a vine. Now in the psychological laboratory, he is confronted with an apparently inaccessible banana. A rope, however, is hanging nearby. *If the animal sees the similarity between the rope and the vine, or between his jungle method and the one now possible, he may solve the problem immediately*" (p. 122). (Italics are the reviewer's.) Now this explanation of the ape's accomplishment in terms of past experience which at first sight seems to be the scientifically simplest explanation actually turns out on closer scrutiny to demand much more in the way of memory and intellectual ability than the proposition that the animal simply sees the relevance of the rope to the banana which is immediately given. This assumption should not be too difficult to make since it was pointed out earlier that the difference between classical (mechanical) and instrumental conditioning lies in the fact that conditioning oc-curs much more easily when in the direction of more relevant re-sponses. Certainly if the principle of relevance is basic to conditioning it may be accepted for the much more complicated case of insightful behavior. The welter of factors involved in acquiring skill and learning could be better ordered and made more meaningfully connected if some structure were seen behind the facts in question.

The lack of an adequate theoretical framework plagues the reader most in the concluding chapter on thinking. Reasoning, we are told, is

implicit trial and error, it is a form of controlled association, it is a combining of past experiences in order to solve problems which cannot be solved by mere reproduction of earlier solutions. At the same time the role of *direction* in reasoning and recall is emphasized, but how this factor operates with trial and error, past associations, and mere reproduction of earlier solutions to problems is not faced. We are here smack up against the problem of organization which many workers in biology as well as psychology realize cannot be dealt with adequately except as a problem in its own right. From the reviewer's experience such problems cannot be evaded even in the beginning course because many students have already faced them in courses in philosophy, logic, biology, and elsewhere.

The section on motivation of behavior seems to this reviewer to be best in the opening chapter dealing with physiological drives such as hunger, thirst, and sex where the material is largely drawn from experimental sources. The chapter on common social motives reads too much like a re-wording of the instinct psychology with too little use made of laboratory findings relevant to the topic. The chapter on conflict opens with sources of conflict in the environment and in the individual and then presents topological representation of conflict situations as an "interesting and illuminating method of representing and analyzing conflict situations" (p. 245). However, in the succeeding treatment of reactions to conflict such as compensation, identification, phantasy, projection, repression, and experimentally produced conflicts there is no further use made of Lewinian concepts. Again the integration must either be made by the instructor or the student will suffer from intellectual indigestion. Other alternatives are to omit topological representation or to put it at the very end of the chapter, pointing out that some, much, or most of the material discussed (depending upon the degree to which the instructor knows topological psychology) can be diagrammed in these terms. The author's penchant for trial and error pops out again in his recommendation of it as a possible solution in the alleviation or cure of conflict, contrary to the usual emphasis on rational procedures in psychotherapy. Since the patient admittedly knows why he is trying various lines of action, namely to find a way out of his conflict, it is doubtful if the procedure recommended is truly trial and error as Munn says.

The section on feeling and emotion which follows the one on motivation of behavior might well have preceded it, as affective states have been regarded by almost everyone as motivators of behavior. The main findings in the field are well covered with one or two exceptions. In the discussion of the Cannon-Bard theory the inhibitory function of the cortex is not mentioned and in the diagram illustrating the contrasting features of this theory as against the James-Lange theory the cortico-thalamic inhibitory path is not even shown. In view of the great im-

portance of the role played by the cortex in inhibiting emotion through positive inhibitory regulation and in allowing emotional expression through release of inhibition, the account offered here is entirely inadequate if not misleading, as reference to Bard's exposition in the *Handbook of General Experimental Psychology*, pp. 305–307, will show. Both text and diagram need this aspect of the theory for a correct as well as complete statement.

The following section, *Knowing Our World*, deals with attention and the special senses. Munn has here done an excellent job of boiling down the classical, and for the most part, stereotyped material and he has made it attractive by the use of well-chosen diagrams and half-tones. In view of the tremendous use made during the war years and now of material from the fields of sensation, perception, and psychophysics, not to mention their interrelations with sensori-motor learning, the time is past when we can rest content with traditional accounts of these fields. There is a wealth of material not yet in any text which modifies the whole approach to sensory processes and bears on every other field of psychology which should form part of the elementary student's equipment. Why recent work in some fields finds its way almost at once into textbooks and equally important work in other fields must wait a generation or more is hard to understand. For example, the explanation given of constancy is naive in the extreme in the light of not-so-recent work. And the epoch-making contributions by Katz find no reflection in the treatment of vision even though they had been in print for 35 years at the time this book appeared.

Several inaccuracies in terminology and fact should be corrected in future editions, such as: brightness and lightness should not be used interchangeably, and unless film and aperture modes are distinguished it is impossible to appreciate their difference; the assumption that Hering "neutral gray" is a constant or a general phenomenon rather than a special case is not tenable in view of work by Koffka and others; the discussion of retinal mixture versus overlapping of lights is so unclear it is impossible to determine what is meant and if it is correct; the usual explanation of *Flor* contrast as due to softening or obliteration of contours is palpably wrong and needs to be supplanted by the correct explanation given by von Bezold in the early part of the present century; the interchangeable use of note and tone, so common in discussions of hearing, should be replaced by more precise terminology in which tone relates to hearing-experience and note to the printed symbol. Only one figure of sensory qualities, the double cone for vision, is given although the smell prism and the taste tetrahedron are just as good in their respective modalities. The value of 1/3 as the Weber fraction for temperature is altogether too large to be representative following Culler's work.

In general the chapters just considered rely too much, on the theoretical side, on past experience and similar explanations and suffer

from a lack of unifying principles by which intra- and inter-sensory material may be related as well as unified with other psychological processes. If, as admitted, principles of organization are effective in attending, are they not perhaps also of importance in perception, and taking a further step, in learning and thinking as well? Recognition of such principles might unify and simplify psychology for the beginner.

The seventh and final section of the book, *Individual Differences*, contains chapters on statistics, intelligence, aptitudes, and personality. The chapter on statistics, kept until this section as an "Introduction to Statistical Analysis of Individual Differences," might well come earlier, especially for use in courses with laboratory. However the chapter can be introduced as it stands in almost any part of the course so its actual position matters little. The other three chapters form a fitting close to the book, entirely in the spirit of the more experimental portions in being packed with concrete material. Intelligence is approached from the historical angle and the important question of heredity and environment is quite fully discussed. The discussion of factor analysis— including the fundamental factors found by Thurstone, and the illustratory material from test batteries—make this chapter unique for an elementary presentation and one of the finest things in the book.

Similarly the chapters on aptitudes and personality are extremely well presented and again demonstrate the author's ability to condense a large amount of fact into a relatively small compass. In the chapter on personality the discussion includes methods of approach such as case history, rating, paper and pencil tests, behavior tests, interviews, free association, dream analysis, and projective methods, and also physique and temperament, role of the endocrines, and abnormal states. The open, empirical treatment here is more acceptable because the subject is more familiar to the average student and personality as a concept already provides some structuration by which its data can be ordered.

Taking the book as a whole, what are its pros and cons? On the plus side it is an excellent text in so far as it provides a wealth of concrete factual data chosen from widely different sources both within and outside psychology proper. With some exceptions it represents present-day scientific psychology very fairly. The student should come aways from this text with respect for the scientific approach not because he has been told that it works best in various fields but because he has found that material obtained by scientific methods can be applied to many different life situations and leads to further fruitful discovery. If the author is unable to accept a theory *in toto* his criticism is so mild and fair that the student's respect for the theory as well as for psychology in general is in no wise diminished. This and the catholicity of Munn's approach should exert a very good effect on coming generations of psychologists. Too often personal or institutional loyalties lead even graduate students to belittle men and work done outside their own bailiwicks with effects

detrimental both to themselves and to psychology. This book should serve as an excellent corrective to this tendency.

On the negative side of an otherwise fine piece of writing and presentation must be noted the lack of an integrating and unifying point of view which has been pointed out in our previous discussion. This lack results in a looser and more disjointed treatment than is necessary in the light of present advances along various fronts. This is not meant to imply that Munn himself does not have a point of view. As we have seen, careful reading reveals that for him trial and error is the great principle operating in human behavior and a number of indications are present that he believes in what has been dubbed an "atomistic logic," *i.e.*, proceeding from "simples" to "complexes." But having brought into his discussion of conditioning the principle of relevance, into thinking the principle of direction, into some sensory experiences primitive organization, and having recognized other whole-properties as well, he is under obligation to apply them more generally where they are applicable or at least to square them with the fundamental principles he believes are operative. Perhaps he has done this and this reviewer has missed it. If so, then it is probable that most students will fail to see how it all goes together.

The tendency to over-simplify has already been pointed out with reference to certain neural diagrams but it occurs much more frequently in the textual discussion where it leads the author, in making points which are quite valid in themselves, to say things he cannot possibly mean as they stand. For example, in writing about the influence of animal experiments on the theoretical basis of psychology, there appears the remarkable statement: "After all, learning is learning and vision is vision whether it occurs in man or animal" (p. 10). But the vision of the most widely used laboratory animal, the white rat, is very different from that of man, from retina to higher cortical centers, and Munn later points out that "Insight is rare in animals, not quite so rare in children, and quite common in human adults" (p. 109), meaning to distinguish among kinds of learning. One finds too many statements like this which take a good deal of explaining to mitigate.

There can be little doubt the present book will set a pattern for future introductory texts. The double columns while providing a shorter reading line and more words per page also make possible wider spaces for illustrative material and marginal notes. The wealth of charts, diagrams, and pictures lessens the instructor's blackboard work and should prove a boon to places where laboratory work cannot be given. In this text psychology appears as a positive, if not positivistic, science. If it were possible to combine what Munn has done with more emphasis on methods and unifying principles we should be much nearer the perfect presentation of present-day psychology everyone desires.

HARRY HELSON.

Bryn Mawr College.

BRIDGES, J. W. *Psychology normal and abnormal.* Toronto: Sir Isaac Pitman & Sons, 1946. Pp. xviii+470.

Except for dropping the chapter on philosophical foundations, the splitting and partial revision of the chapter on reflexes and instincts, and the annotation of the extensive bibliography, the 1946 edition has "identical twin" resemblance to its 1930 predecessor. The appearance of the revision does, however, call attention again to a book with a classical timelessness of integration (despite an eclectic tolerance), stimulating hypotheses, and a style reminiscent of William James. The general reader and even the professional psychologist wili find interest and value here, although the book was designed for the introductory psychology course of pre-medical and medical students.

Bridges chooses to give the distilled essence of a topic rather than to lead the reader to a conclusion from the raw data of experiments and case studies. The few graphs, tables, and other reference to specific studies are illustrative only. Just as the dramatist's words furnished the bare Elizabethan stage, the emphasis on the logic of the argument in this book seems to stimulate more associations and imagery than one gets from many texts replete with illustrations. Many more of the quotations are from English and French psychologists than one meets in most American texts.

The chapter headings might have come from any of a number of general psychologies, but the plan of devoting the first half of each chapter to normal behavior and the remainder to the related abnormalities makes a distinctive pattern throughout the book. Technical words are italicized and well defined. The chapter on applied psychology delimits that field with such precision and perspective that it ought to be widely reprinted.

An error not corrected from the 1930 edition is the taking of the standard deviation from the median. Emphasis on the older studies, e.g., Downey will-temperament tests, is heavier than on those of the last two decades.

The problem of how to teach psychology in medical schools or to pre-medical students seems to have led in at least three main directions. Some have emphasized a sociological-psychological approach, as in Pressey's *Life,* since this is a common medical blind spot. Others have stressed the genetic-psychosomatic attitude, as found in such authors as Maslow and Mittlemann. A third group believe that medical students are more highly motivated and gain more insight from the contrasts and comparisons of the normal and abnormal. Bridges from his experience as the first professor of abnormal psychology on a medical faculty has provided an effective text for the last group.

GEORGE M. HASLERUD.

University of New Hampshire.

GRAY, J. S. *Psychology in human affairs*. New York: McGraw-Hill, 1946. Pp. viii+646.

While this book is, in many respects, a successor to Gray's previously edited text, *Psychology in Use* (American Book Co., 1941), in that it discusses the applications of psychology to the main fields of practical life, and represents the co-authorship of eleven other contributors, nevertheless it is not merely a revision. With two exceptions, the co-authors are new. They are less well known than those of the former book, but Gray has himself taken a more active part in the actual writing of the text. Several chapters appear for the first time, such as "Psychology in Speech Correction," "Psychology in Music, Art, and Literature," and "Psychology in Military Affairs." Others appear under new titles, and are written from a new viewpoint.

Perhaps the outstanding characteristic of the book is its emphasis on factual material. For example, Chapter II, on "Psychology in College Life" contains twenty tables and four graphs. Chapter III, on "Child Development" contains ten graphs and twenty-one tables. Much of this material is new to textbooks, and with few exceptions, the references are to studies published after 1930. The general effect of this emphasis on experimental data and practical findings is to require a change in teaching methods on the instructor's part. His function is no longer to supplement the text with up-to-date illustrative material, but rather to interpret and evaluate that which is given. Less supplementary assigned reading is needed, and much more digesting of the text by the student. The art of reading and interpreting tables and graphs is one which requires special training. Many students are allergic to statistics, though this is not in itself an argument for using them sparingly, if the instructor is competent to vitalize them. But the chief value of facts is to illustrate and support laws, principles and theoretical formulations. They are most effectively used in the inductive development of a topic. Most of these facts are and should be promptly forgotten by the student, so that his memory is freed for the permanent retention of the principles. The immature student needs much expert guidance in recognizing the bare essentials of fact to be learned. While this book is many strides ahead of the type which presents only unsupported assertions, or illustrations selected for their patness only, does it err slightly in the opposite direction?

Another important characteristic of the book is its emphasis on the practical. This is to be expected in an applied psychology text, but is seldom achieved. Omnibus books too often give an impression of sketchy remoteness, with little practical contact, while technical treatises are written for the advanced student who wants specialized information. This book, by achieving a compromise between these two extremes and by emphasizing the practical aspects of each field for the layman, fills a

real need. Its range of topics is wider, its treatment more complete than is customary in such books.

In spite of its up-to-dateness, the book occasionally presents old data which have been superseded, or theory which is now modified. In one or two cases, quite erroneous statements appear, such as the following, in connection with a discussion of the topic of I.Q. constancy, on pages 91–92:

> If the child develops mentally at exactly the same rate as other children tested, his I.Q. will remain constant. However, if his mental development is faster than that of other children, his I.Q. will increase. Likewise, if his mental development is slower than that of other children, his I.Q. will decrease.

The author seems to have confused *constancy* of I.Q. with *normality* of I.Q., for if the statement were taken as it stands, it would mean that no child's I.Q. is constant if he has a faster or slower developmental rate than the average child.

An innovation in this book which probably has pedagogical value and would be used oftener by authors of texts if publishers would let them, is the table of contents at the beginning of each chapter. The addition of page numbers would increase its value. A word must be said in criticism of certain reproduced charts, in which the reduction in size of print needed to get them on the page has made them unreadable; for example, those on pp. 474 and 573. Otherwise the style of the book is good.

Many psychology teachers will welcome this book either as a supplement to the general course in psychology, or as a second course to follow the introductory one. Those who teach adult extension classes will find it an excellent survey text, both meaty and comprehensive.

<div align="right">ARTHUR G. BILLS.</div>

University of Cincinnati.

LUCK, J. M., & HALL, V. E. (Eds.). *Annual review of physiology* (Vols. VII & VIII). Stanford Univ. P.O.: Annual Reviews, Inc. and American Physiological Society, 1945, 1946. Pp. vi+774 (Vol. VII). Pp. vi+658 (Vol. VIII).

These are Volumes VII and VIII of the annual series begun in 1937 and published jointly by the American Physiological Society and Annual Reviews. It is the declared editorial policy of the *Review* that "encouragement is given only to preparation of reviews which survey the important contributions of the preceding year or biennium, which appraise them critically, and evaluate with discrimination the present status of the subject. Comprehensive reviews in which the task of the author is one of compilation rather than of appraisal are deliberately eschewed." Despite this policy, some of the reviews are principally compilations or annotated bibliographies. And some are rather spotty

compilations mixed with evaluation, while only a few reach the goal of really critical reviews of the literature. On the whole, however, the reader can obtain a picture of the more significant aspects of research progress in the respective fields covered by the review.

Volume VII contains 26 chapters written by 30 authors, and Volume VIII contains 25 chapters by a total of 29 authors. In each case, bibliographies of literature cited in the various reviews total about 4,000 references. At the end of each volume is an author index of about 4,000 items and a subject index of about 40 pages in length.

Because the reviews are written by physiologists for physiologists, more than half of each volume is of no interest to the psychologist except that occasionally there is a brief treatment in a sentence or two of some psycho-physiological problem. The psychologist, however, who wants to find out what has happened recently in some special aspect of physiology relevant to his field of teaching or research will find that his best bet is to go to these volumes and to consult the excellent author and subject indices before resorting to other less up-to-date textbooks or to more laborious methods of library research.

More than that, however, many of the special chapters in physiology are good reading for psychologists engaged in the respectively related field of psychology: for genetic psychology, *Physiological aspects of genetics* VII and VIII) and *Developmental physiology* (VII and VIII); for sensory psychology, *The special senses* (VII) and *Audition* (VIII); for neural mechanisms of behavior, *Electrical activity of the brain* (VII), *Conduction and synaptic transmission in the nervous system* (VII), *Nerve and synaptic conduction* (VIII), *The visceral functions of the nervous system* (VII and VIII); and for a general review of physiological psychology, *Physiological psychology* (VII and VIII).

The contrast between the chapters on physiological psychology in the two volumes deserves a special comment. In Volume VII, Stone has presented a careful and critical review. He has covered thoroughly the recent literature and has appraised its strength and shortcomings so that the reader can see what has happened and what it means for physiological psychology.

The same chapter in Volume VIII by Seashore, however, does neither of these two things. It opens with a philosophical discussion of the mind-body problem and of the scientific approach to it. The chapter then proceeds to summarize the present status of individual differences in skills, abilities, aptitudes and capacities. Finally, it gives a general summary of the effects of extreme working conditions upon the effectiveness of human performance. Thus Seashore's chapter spends a lot of time on problems which are not physiological psychology, in any reasonable definition of the field; and he fails to review or appraise the recent literature in the field. As a consequence, the physiologist reading the two chapters is likely to be bewildered by two so very different concepts of physiological psychology.

Looked at in perspective, these two volumes of *Annual Review of Physiology*, like previous volumes in the series, are an extremely valuable aid, not only to physiologists, but to all those for whom physiology is an important ancillary subject. By and large, the chapters give scholarly up-to-date appraisals of their respective fields. As he has stated before, this reviewer feels that a companion volume, giving an annual review of psychology, would be an invaluable aid to psychologists, which would help us "keep up with the literature" and give us better perspective on the developments in our field.

CLIFFORD T. MORGAN.

The Johns Hopkins University.

BARKER, ROGER G., WRIGHT, BEATRICE A., AND GONICK, MOLLIE R. *Adjustment to physical handicap and illness: a survey of the social psychology of physique and disability.* New York: Social Science Research Council, 1946. (Bulletin No. 55.) Pp. xi+372.

This is another in the excellent series of research summaries sponsored by the Social Science Research Council. The authors have earned special commendation by providing intelligently critical comments as to the assumptions and thinking of earlier investigators, rather than merely reporting data and conclusions; by writing this summary of prior research into a theoretical frame of reference (topological psychology); and by introducing some well-chosen original material to illuminate the inferences drawn from published sources.

From their survey the authors have eliminated somato-psychological studies of age, sex, race, and speech defects, on the ground that these have recently been covered adequately by other reviewers. Leprosy is discarded as a minor problem in the western world. Of the remaining areas, detailed reports are presented on: normal variations in physique; crippling; tubercular conditions; impaired hearing; and acute illness. Bibliographies are added on: visual disability; cardiac conditions; diabetes mellitus; cosmetic defect; rheumatism; and cancer.

The least satisfactory chapter is that on variations in normal physique. There is a good section on size changes at adolescence, but the discussion of variations in adult size is rather elementary, and no mention whatever is made of Sheldon's *The varieties of human physique* and *The varieties of temperament.* Even if one does not accept Sheldon's theory, his work can hardly be ignored. The authors occasionally dip a hesitant toe into the cold waters of endocrinology, genetics and autonomic nervous function, then withdraw hastily. Clarity would have been improved by frankly excluding such material.

Outstanding treatments of crippling, tubercular conditions and impaired hearing more than atone for any shortcomings of the earlier chapter. Particularly interesting is the mode of analysis in terms of *overlapping situations.* A disabled person is able to function on a par with normals in some environments; he is decisively barred from other

situations; but between these extremes will fall a range of ambigu us conditions in which the handicapped person may participate, but und r difficulties. The Lewinian concepts of barrier, valence, potency and congruence are used fruitfully to show basic similarities between the situations facing the orthopedic cripple, the tubercular, the deaf and the individual with acute illness.

The person with impaired hearing, for example, functions in many situations unnoticed by his normal associates. If the behavior involved does not require auditory controls, he may compete on equal terms. Where hearing is involved, he may be handicapped and subject to extra criticism, since his impairment is not obvious and many normals (e.g., school teachers) fail to make allowances for it. The barriers in his field are indefinite (as contrasted with the orthopedic cripple, for example, to whom certain activities are plainly impossible), and this condition often gives rise to vacillation and instability. The valence of full-normal ac- tivity is positive and high, but the valence of failure and criticism is negative and high. Thus physically handicapped persons are likely to show the familiar symptoms of conflict.

We occasionally feel, in these topological analyses, that there is an unstated shift from the topology of the external situation (geographical environment) to the situation as perceived by the individual (behavioral environment). In the case of cripples, for example, some activities are objectively impossible, whereas others are subjectively considered to be impossible. It is clear that these two should not be treated as identical, and yet that impression is sometimes given. If the entire analysis were erected on a perceptual basis, this uncertainty could have been avoided.

The importance of the individual's perception of his defect, and of his behavioral field, is well illustrated by the discussion of family attitudes toward the handicapped. Many parents reject the handi- capped child, while others over-protect him and keep him in an infantile status. Optimum adjustment seems to be achieved when the parents adopt an understanding, objective attitude which focuses the child's attention on realistic assessment of the situation. Excessive sympathy and pity are likely to encourage exaggeration of barriers and exclusion of many possibilities for normal participation.

The final chapter on employment of disabled persons gives a realis- tic and well-considered treatment of this problem, the solution of which is basic to optimum adjustment of handicapped adults.

ROSS STAGNER.

Dartmouth College.

LEWIS, CLAUDIA. *Children of the Cumberland.* New York: Columbia Univ. Press, 1946. Pp. xviii+217.

Before going to the Southern highlands, Miss Lewis was a teacher in the Harriet Johnson Nursery School in Greenwich Village, New York

City. She compares the behavior of the children in the nursery school which she established in the mountains of Tennessee with the behavior of her Greenwich Village pupils. The majority of the material presented in the book concerns the mountaineer subjects.

A considerable part of the volume consists of a collection of incidents involving child care or child behavior. These range in length from single sentences to a page or more, in age of subjects from birth to senility, in form from dialogue to descriptive essays. They are extremely readable and serve to render very vivid the life of the mountain people.

Miss Lewis does not attempt to present quantitative measures, but for this she cannot be reprimanded. It is apparent that she devoted very full days to the nursery school, and that research had a secondary place. Nevertheless, her thinking is quantitative. She emphasizes the diversity of individual behavior which takes place in both schools, and makes clear that there is overlapping between the schools. However, it is her belief that there are large differences in central tendencies between her two kinds of subjects. With this contention, it is likely that nearly every person who is familiar with both types of children will agree.

Miss Lewis finds more spontaneity, more energy, more conflict, more aggression in the New York group. The mountain children are more placid, more compliant, more quiet, and in some respects, better adjusted.

She does not find the explanation of these differences in any one factor. Among the probable causes mentioned are the following: the differences in spaciousness of the environment, differences in climate, health, and nutrition, differences in sleeping habits, differences in infant care and family structure, differences in discipline, and differences in environmental stimulation.

Throughout the book, Miss Lewis shows an excellent understanding of child development in both New York and Tennessee—not an easy achievement. She also displays a high degree of ability as a writer. This combination of traits makes her book one of the best in the "child in a culture" field. It should be profitable alike to teacher, parent, psychologist, and sociologist.

The reader may sample for himself some of Miss Lewis's attitudes and style in the following quotation from her concluding chapter:

No, now that this study is made I am not packing my trunks with the intent of moving down to Tennessee, building me a cabin, taking to the "simple life" and rearing my hypothetical children in the way the Summerville families do. For us it is not a question of attempting to turn the clock back in that way, which, indeed, would be as impossible as it would be undesirable. It is rather a question of trying to bring to Greenwich Village a little more of . . . Summerville life . . . .

WAYNE DENNIS.

University of Pittsburgh.

LEEPER, ROBERT. *Psychology of personality.* Ann Arbor: J. W. Edwards, 1946. Pp. 167.

The format of this book, resembling that of a laboratory manual or workbook, may lead many to overlook this significant treatment of personality and mental hygiene.

The author's organization apparently proceeds from an assumption which has increasingly impressed itself on the reviewer in recent years: namely, that, in order to have functional significance, any treatment of mental hygiene must be based on a consistent theory of personality processes. Moreover, such a treatment should not be left among the author's unverbalized potentialities, but should be given systematic formulation. It is to this task that the major emphasis of Leeper's book is devoted.

Leeper's treatment is mainly an elaboration of the following thesis:

In general, the term "personality" covers three things: (1) the person's motives, and especially his emotional motives, or ways in which he responds emotionally in different life situations; (2) the general techniques by which, characteristically, he tries to attain satisfaction for these motives; and (3) the background of meanings or pictures of reality which determine the motives and types of adjustive responses of the person (p. 5).

The discussion of motivation distinguishes between physiological and emotional motives and between positive can negative emotional motives. While the latter distinction seems arbitrary, since the "negative" motives can be regarded as the products of frustration of the "positive" motives, it serves a useful expository purpose when the author deals with motivational differences existing between well adjusted and poorly adjusted personalities.

The techniques by which the person tries to attain satisfaction for his motives are considered from two points of view: (1) the nature of the learning processes, and (2) the description of effectual and ineffectual adjustment techniques. The learning processes are treated with due attention to the dynamic complexities recognized in modern learning theory. In addition to describing the usual techniques employed by maladjusted personalities, the author discusses some of the major techniques by which superior personalities distinguish themselves.

In his discussion of the "background of meanings," the author deals competently with an aspect of behavior which seldom receives the emphasis merited by its significance for personality dynamics. Leeper's thesis is that " . . . a person cannot govern his behavior just by what is objectively and actually true, but . . . must forever live and react in terms of properties which he infers as existing because of his experience in previous situations" (p. 92). This view that behavior is determined not as much by the character of objective reality as by the individual's interpretation of reality (through the phenomenon of emotional transference) is supported in terms of the principle of equivalence of stimuli and the principle of substitute response or displacement. Treated in

these terms, the "background of meanings" is seen to be an aspect of personality whose importance has been emphasized by such widely separated disciplines as the research of animal psychologists and the clinical observations of psychoanalysts.

It is the reviewer's opinion that through the medium of Leeper's book the principles of personality functioning are made understandable to the average undergraduate without undue simplification or loss in organic quality. For this reason, it seems regrettable that the book was not produced in a form more likely to have wide distribution.

The book will probably be disappointing to those who feel that a textbook should serve as a compendium of psychological research findings. While the author draws freely upon research sources, these tend to lose their distinctive identities in the author's discussion. No bibliography or index is provided.

BERT R. SAPPENFIELD.

Montana State University.

KELLEY, DOUGLAS M. *22 cells in Nuremberg.* New York: Greenberg, 1947. Pp. 245. $3.00.

The author of this book was for five months the official psychiatrist at the Nuremberg prison and in that capacity made psychiatric examinations af all the 22 top-ranking Nazi prisoners. The customary medical and psychiatric procedures were supplemented by Rorschach personality tests and Wechsler-Bellevue intelligence tests given by the author's fellow officer, Capt. G. M. Gilbert. The examinations and tests were further supplemented by information obtained from former intimate associates of the accused and from motion pictures, speeches, writings, and other records.

Except for Rudolf Hess and Hermann Goering, this was the first psychiatric study to be made of any of the accused. In view of the fact that twelve of the group are no longer living and that all but three of the others were disposed of by prison sentences of from ten years to life, the documents and conclusions of Dr. Kelley are destined to be of lasting historical interest.

The task which the author set himself was not merely or chiefly to determine the degree of mental responsibility of the subjects, but rather to investigate their basic personality patterns. He wanted to find out what these men were like who had made themselves masters of eighty million people, and what factors in childhood, youth, and later years had made them what they were. The book attempts to answer these questions in language sufficiently nontechnical to be intelligible to readers who are relatively unfamiliar with esoteric theories of personality. We are informed that more detailed reports of the work will be published later in professional journals, and that transcripts of the interviews and other records will ultimately be available to historians.

Among those to whom most space is devoted are Goering (27 pages), Hess (22 pages), Ley (21 pages), Rosenberg (13 pages), and Streicher

(11 pages). The other 17 members of the group get from three to eight pages each. By good luck the examination of Ley had been completed before he committed suicide, and we are told that the post-mortem examination of his brain confirmed the psychiatric diagnosis.

With the exception of one chapter, the book is concerned entirely with the 22 Nazis who were studied first-hand by the author. The additional chapter presents a 35-page portrait of Hitler based on information and comments obtained from the Führer's contemporaries, his aides, his personal physicians, and his secretaries. Some of this information is new, and the author's interpretations differ in several important respects from those which have been current.

Within the limits of a brief review it is not possible to summarize the author's interpretations of the individual subjects. In fact, each portrait as sketched is a unified gestalt that almost defies further condensation. The sitters for these portraits composed indeed a motley group. They ranged from the stupid to the highly intelligent; from the semi-insane to the stable and well integrated; from the shrewd and talented leader to the errand-boy hanger-on seeking in Hitler a father surrogate. But there were three characteristics which they had in common: inordinate ambition, debased ethical standards, and a hyperdeveloped nationalism that justified anything done in the name of Germandom— plus, of course, an economic and political environment that allowed full play to their ruthless wills.

The author's conclusion is that Nazism was a "socio-cultural disease," epidemic among our enemies but endemic everywhere. He tells us that the Nazi leaders were not the rare and spectacular types that can be expected to appear only once in a century. Instead, neurotics like Hitler, with "hysterical disorders and obsessive complaints, can be found in any psychiatric clinic." Similar ones, thwarted and discouraged, but determined to do great deeds, roam the streets of every American city. "Strong, dominant, aggressive, and egocentric personalities like Goering . . . can be found anywhere—behind big desks deciding big affairs as businessmen, politicians, and racketeers." We hardly need to be reminded that men strongly resembling some of these types occasionally win election to our highest law-making bodies or to the governorship of a great state.

Dr. Kelley has analyzed for us 22 types of totalitarian-virus, has described the soil in which they thrive, and has indicated some of the means by which society can protect itself against them. His book will inevitably be compared with one written by another psychiatrist— Brickner's *Is Germany Incurable?* Of the two, the reviewer finds Kelley's less controversial and no less challenging.*

LEWIS M. TERMAN.

Stanford University.

* Since this review was written, *Nuremberg Diary*, by Capt. G. M. Gilbert, has been published. This book should be read along with Dr. Kelley's. L.M.T.

BOOKS AND MATERIALS RECEIVED

ALLPORT, G. W., AND POSTMAN, LEO. *The psychology of rumor*. New York: Henry Holt, 1947. Pp. vii+247.

ALSCHULER, ROSE H., AND HATTWICK, LA BERTA W. *Painting and personality*. (2 Vols.) Chicago: Univ. of Chicago Press, 1947. Pp. xi+590.

AXLINE, VIRGINIA MAE. *Play therapy*. Boston: Houghton Mifflin, 1947. Pp. xii+379. $3.50.

BARUCH, DOROTHY, AND MONTGOMERY, ELIZABETH. *The girl next door*. Chicago: Scott, Foresman, 1947. Pp. 256.

BAUMGARTEN-TRAMER, F. *Der Rorschach—Test im Lichte der Experimentellen Psychologie*. Archivio di Psicologia Neurologia e Psichiatria, 1946, 8, No. 2. Pp. 37.

BECK, H. P. *Men who control our universities*. New York: King's Crown Press, 1947. Pp. x+229.

BOWERMAN, WALTER G. *Studies in genius*. New York: Philosophical Library, 1947, Pp. 343.

CARTER, L. F. (ED.) *Psychological research on navigator training*. Army Air Forces Aviation Psychology Program, Research Reports. Report No. 10. Washington: U. S. Government Printing Office, 1947. Pp. iii+186.

COLE, LUELLA, AND MORGAN, J. J. B. *Psychology of childhood and adolescence*. New York: Rinehart, 1947. Pp. xi+416.

COUNT, E. W. *Brain and body weight in man: their antecedents in growth and evolution*. Annals of the New York Academy of Sciences, 1947, Vol. XLVI, Art. 10. Pp. 993–1122.

COWLES, E. S. *Don't be afraid*. Chicago: Wilcox & Follett, 1947. Pp. xv+254.

DAVIS, F. B. (ED.) *The A. A. F. qualifying examination*. Army Air Forces Aviation Psychology Program, Research Reports. Report No. 6. Washington: U. S. Government Printing Office, 1947. Pp. iii+266.

DAVIS, F. B. *Utilizing human talent*. Washington: American Council on Education, 1947. Pp. ix+85.

DICHTER, E. *The psychology of everyday living*. New York: Barnes & Noble, 1947. Pp. x+239.

ERICKSON, C. E. (ED.) *A basic text for guidance workers*. New York: Prentice-Hall, 1947. Pp. x+566.

FALVEY, HAL. *Ten seconds that will change your life*. Chicago: Wilcox & Follett, 1947. Pp. 96.

FIELD, HARRY. *Midiendo la Opinion Publica*. Mexico, D. F.: In-

485

stituto de Estudios de Psicologia Social y Opinion Publica, 1945. Pp. 56.

FORD, MARY. *The application of the Rorschach test to young children.* Institute of Child Welfare Monograph, No. 23. Minneapolis: Univ. of Minnesota Press, 1946. Pp. xii+114.

FREDERICK, R. W., KITCHEN, P. C., AND McELWEE, AGNES R. *A guide to college study.* New York:Appleton-Century, 1947. Pp. viii+341.

GARRETT, H. E. *Statistics in psychology and education.* (3rd Ed.) New York: Longmans, Green, 1947. Pp. xii+465.

GILBERT, G. M. *Nuremberg diary.* New York: Farrar, Straus, 1947. Pp. 471.

HAGAN, W. A., *et al. The relation of diseases in the lower animals to human welfare.* Annals of the New York Academy of Sciences, 1947, Vol. XLVIII, Art. 6. Pp. 351–576.

HALL, R. B. *Area studies: with special reference to their implications for research in the social sciences.* New York: Social Science Research Council, 1947. Pp. iii+90.

HALL, V. E. (Ed.) *Annual review of physiology* (Vol. IX, 1947). Stanford Univ. P.O.: Annual Reviews, Inc. & American Physiological Society, 1947. Pp. vii+736.

JELLINEK, E. M. *Recent trends in alcoholism and in alcohol consumption.* New Haven: Hillhouse Press, 1947. Pp. 42.

KITAY, P. M. *Radicalism and conservatism toward conventional religion. A psychological study based on a group of Jewish college students.* Teachers College, Columbia Univ., Contr. to Educ., No. 919. New York: Bureau of Publications, Teachers College, Columbia Univ. 1947. Pp. vii+117.

KRACAUER, S. *From Caligari to Hitler.* Princeton: Princeton Univ. Press, 1947. Pp. xi+361.

LE CRON, L. M., AND BORDEAUX, JEAN, *Hypnotism today.* New York: Grune & Stratton, 1947. Pp. v+278.

LEIGHTON, DOROTHEA, AND KLUCKHOHN, CLYDE. *Children of the people.* Cambridge: Harvard Univ. Press, 1947. Pp. viii+277.

LEPLEY, W. M. *Psychological research in the theaters of war.* Army Air Forces Aviation Psychology Program, Research Reports. Report No. 17. Washington: U. S. Government Printing Office, 1947. Pp. iii+201.

LINDNER, R. M., AND SELIGER, R. V. *Handbook of correctional psychology.* New York: Philosophical Library, 1947. Pp. 691.

LINK, H. C. *The rediscovery of morals.* New York: E. P. Dutton, 1947. Pp. 223.

LONDON, L. S. *Libido and delusion.* (2nd Ed.) Washington: Mental Therapy Publ., 1946. Pp. xi+259. $3.50.

LOUTTIT, C. M. *Clinical psychology.* (Rev. Ed.) New York: Harper, 1947. Pp. xviii+661.

LUNDBERG, G. A. *Can science save us?* New York: Longmans, Green, 1947. Pp. 122.

LUNEBERG, R. K. *Mathematical analysis of binocular vision.* Princeton: Princeton Univ. Press, 1947. Pp. vi+104. $2.50.

MICHOTTE, A. *La perception de la causalité. Études de Psychologie,* 1946. Vol. VI. Pp. viii+296.

NACHMANSOHN, D., *et al. The physico-chemical mechanism of nerve activity.* Annals of the New York Academy of Sciences, Vol. XLVII, Art. 4. Pp. 375–600.

NICOLE, J. E. *Psychopathology. A survey of modern approaches.* (4th Ed.) Baltimore: Williams & Wilkins, 1946. Pp. vii+268. $4.75.

RADKE, MARIAN J. *The relation of parental authority to children's behavior and attitudes.* Institute of Child Welfare Monograph, No. 22. Minneapolis: Univ. of Minnesota Press, 1946. Pp. x+123.

ROBINSON, F. P. *Effective study.* New York: Harper, 1946. Pp. ix+262.

SADLER, W. *Mental mischief and emotional conflicts.* St. Louis: C. V. Mosby, 1947. Pp. 396.

SANDOW, A., *et al. Muscular contraction.* Annals of the New York Academy of Sciences, 1947, Vol. XLVII, Art. 6. Pp. 665-930.

SHERIF, MUZAFER, AND CANTRIL, HADLEY. *The psychology of ego-involvements.* New York: John Wiley, 1947. Pp. v+525.

SNYDER, W. U. (ED.) *Casebook of non-directive counseling.* Boston: Houghton Mifflin, 1947. Pp. viii+339. $3.00.

SOROKIN, P. A. *Society ,culture and personality: their structure and dynamics.* New York: Harper, 1947. Pp. v+742.

THORNDIKE, R. L. (ED.) *Research problems and techniques.* Army Air Forces Aviation Psychology Program, Research Reports. Report No. 3. Washington: U. S. Government Printing Office, 1947. Pp. viii+163.

THURSTONE, L. L. *Multiple-factor analysis.* Chicago: Univ. of Chicago Press, 1947. Pp. xix+535.

WOLFF, WERNER. *What is psychology?* New York: Grune & Stratton, 1947. Pp. vii+410.

WOODWORTH, R. S., AND MARQUIS, D. G. *Psychology.* (5th Ed.) New York: Henry Holt, 1947. Pp. iii+677.

WORTIS, S. B., *et al. Physiological and psychological factors in sex behavior.* Annals of the New York Academy of Sciences, 1947, Vol. XLVII, Art. 5. Pp. 603-664.

L'Anneé Psychologique. Henri Piéron (Ed.). Forty-third and forty-fourth years (1942-1943). Paris: Presses Universitaires de France, 1947. Pp. 856.

Archivio di Scienza della Cerebrazione e dei Psichismi. M. Levi—Bianchini (Ed.). Vol. I-II, 1944-45. Pp. 105-208.

1946 fall testing program in independent schools and supplementary studies. Educational Records Bulletin, No. 47. New York: Educational Records Bureau, 1947. Pp. x+58.

Revista de Psicologia General y Aplicada. Vol. 1, No. 1. Madrid: Instituto Nacional de Psicotecnia, 1946. Pp. 316.

Revista do Centro Psiquiatrico National. Vol. I, No. 1. Rio de Janeiro, Brazil: Imprensa Nacional, 1946. Pp. 161.

Theoria. Ake Petzall (Ed.). Vol. XII, Parts I-II. Lund, Sweden: C. W. K. Gleerup, 1946. Pp. 133.